



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba

REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

**Predicción de efectos aleatorios en modelos mixtos.
Aplicación al proceso de innovación en el sector
manufacturero de la provincia de Córdoba.**

Verónica Arias

Tesis en Maestría en Estadística Aplicada. Escuela de Graduados. Facultad de Ciencias
Económicas. Universidad Nacional de Córdoba, 2019. Córdoba, Argentina



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

UNIVERSIDAD NACIONAL DE CÓRDOBA



**Predicción de efectos aleatorios en modelos mixtos.
Aplicación al proceso de innovación en el sector
manufacturero de la provincia de Córdoba.**

Tesis para optar por el grado de
Magister en Estadística Aplicada

Lic. Verónica Arias

2019

UNIVERSIDAD NACIONAL DE CÓRDOBA



COMISION ASESORA DE TESIS

Directora: Dra. Norma Patricia Caro

Miembros: Dra. Cecilia Inés Bruno

Mgter. Regina Ivana Cavallin

Dr. Fernando García

Fecha de aprobación de tesis

18 de diciembre de 2019

Agradecimientos

A mi familia por su amor, su comprensión en las ausencias y por ser parte de esto tanto como yo.

A Patricia Caro por su paciencia, dedicación y por animarme día a día para que este trabajo llegue con éxito a su fin.

A Viviana Giampaoli por su inmensa generosidad.

A mis amigos y compañeros de trabajo, por su apoyo y por hacerme disfrutar de la tarea de todos los días.

A la Facultad de Cs. Económicas UNC y a la Dirección de Estadística y Censos de la Provincia de Córdoba, por ser parte fundamental en mi formación y por darme el espacio para el desarrollo profesional.

A los miembros de la comisión asesora de tesis, Dra. Cecilia Inés Bruno, Mgter. Regina Ivana Cavallin y Dr. Fernando García, por dedicar su valioso tiempo a la revisión de este trabajo.

Palabras clave

MODELO LOGISTICO MIXTO – EFECTOS ALEATORIOS - INNOVACIÓN EN LA INDUSTRIA – CÓRDOBA ARGENTINA

Resumen

Los principales objetivos de la presente investigación son, en primer término, determinar los factores asociados a la innovación en productos y/o procesos en empresas industriales de la provincia de Córdoba y, en segundo término, en el marco de un modelo logístico con efectos aleatorios, mejorar la predicción de la variable respuesta en empresas que no formaron parte del ajuste del modelo.

Para modelar la probabilidad de obtener resultados innovadores en productos y/o procesos mediante la consideración de un conjunto de factores característicos de las empresas y de su entorno se utiliza un modelo logístico mixto. Esto permite incorporar al análisis las mediciones presentadas por cada empresa a través del tiempo, siendo éste un importante aporte de la presente investigación.

Los principales resultados de este trabajo muestran que la probabilidad de obtener resultados innovadores exitosos en producto y/o procesos se ve influida por factores como el tamaño de la firma, el porcentaje de las ventas destinados a inversión en actividades de innovación, la continuidad en el esfuerzo innovador, la utilización de fondos de programas públicos de fomento y la vinculación con otras empresas e instituciones del sistema nacional de innovación.

En los modelos de efectos mixtos, la predicción de la variable respuesta en nuevas unidades se complejiza. Además del efecto fijo, el modelo cuenta con los efectos aleatorios estimados para cada empresa de la muestra con la que se ajustó el modelo. Avanzando

en este sentido, se realizan estimaciones de los efectos aleatorios para las nuevas empresas a través de cuatro métodos alternativos: Naive, EBP, Regresión Lineal y Vecino más Cercano. Si bien las diferencias entre un método y otro sobre el porcentaje de empresas clasificadas correctamente son pequeñas, la estimación de los efectos aleatorios permitió mejorar la capacidad predictiva del modelo sobre nuevas unidades. De esta manera se levanta una limitación que tenían los modelos mixtos en cuanto a la estimación y predicción de la variable respuesta para una nueva observación, en este caso en la probabilidad de obtener resultados innovadores en productos y/o procesos.

Prediction of random effects in mixed models. Application to the innovation process in the manufacturing sector of Córdoba, Argentina.

Key words

MIXED LOGISTIC MODEL - RANDOM EFFECTS - INNOVATION IN THE INDUSTRY - CÓRDOBA ARGENTINA

Summary

The main aims of the following research are not only to determine the factors related to the innovation of products and/or processes in industrial companies of Cordoba province, but also to improve the prediction within companies that were not part of the model adjustment, considering the logistic model with random effects.

In order to obtain innovative results in products and/or processes, considering particular factors from the companies and their environments, a mixed logistic model is used. This allows the incorporation of the measurements to the analysis, presented by each company throughout time. This is what in the end an important contribution of the following investigation.

The main results of this research show that the probability of getting innovative and successful results within products and/or processes is influenced by different factors such as the size of the enterprise, the percentages of the sales invested in innovative activities, the continuity of the innovative effort, the use of public programs funds, and the connection with other companies and institutions from the innovative national system.

Within the mixed effect models, the variable of predictions in new units is more complex. Apart from the fixed effect, the model has the estimated random effects for each company of the sample from which the model was adjusted. Therefore, it is estimated the random effects for new companies through four alternative methods: Naive, Empirical Best Predictor (EBP), Linear Regression, and Nearest Neighbor. When talking about new units, although the differences between one method and the other, as regards the classified companies, are little, the estimation of random effects allowed the improvement of the predictive quality of the model. As regards the estimating and predicting the response variable for a new observation, a limitation of mixed models is stopped.

Índice

Capítulo 1: Introducción	14
1.1 Definición y planteo del problema	14
1.2 Preguntas de investigación	16
1.3 Objetivos	17
1.4 Aportes del trabajo	18
Capítulo 2: Marco teórico	20
2.1 Modelo lineal	20
2.2 Modelo evolucionista	22
2.3 Sistemas de innovación	25
2.4 Estudios empíricos recientes	30
Capítulo 3: Datos y Metodología	37
3.1 Fuente de datos: características de la encuesta	37
3.2 Muestra y variables	39
3.3. Método	41
3.3.1 Modelos para respuesta no Normal	42
3.3.1.1 Modelo logístico	44
3.3.1.2 Modelo logístico mixto	45
3.3.2 Predicción de la variable respuesta en un modelo logístico mixto	50
3.3.2.1 Naive	51
3.3.2.2 Método del “mejor predictor empírico” (EBP)	51
3.3.2.3 Método de regresión lineal	53
3.3.2.4 Método del vecino más cercano	54
3.3.2.5 Comparación de los métodos de predicción: estudios empíricos	55

3.3.3 Medidas de desempeño del modelo	57
Capítulo 4: Resultados.....	61
4.1 Innovación en la provincia de Córdoba.....	61
4.2 Descripción de la muestra	63
4.3 Determinantes de la Innovación	69
4.3.1 Resultados de la estimación	69
4.3.2 Predicción para nuevos grupos.....	76
Capítulo 5: Conclusiones	80
5.1 Reflexiones finales y contribuciones del trabajo	80
5.2 Limitaciones del trabajo y futuras líneas de investigación.....	85
Referencias Bibliográficas	86
Anexos	91
A1 Cuestionario Encuesta de Innovación y Conducta Tecnológica Córdoba.....	91
A2 Sintaxis procesamiento R	105

Capítulo 1: Introducción

1.1 Definición y planteo del problema

La innovación es un proceso complejo a través del cual las empresas transforman conocimientos en valor agregado. Existen diversas definiciones de innovación. Según el Manual de Oslo (OECD, 2005) se entiende por innovación a la introducción de un nuevo o significativamente mejorado producto (bien o servicio); de un proceso; de un nuevo método de comercialización o de un nuevo método de organización; en las prácticas internas a la empresa, la organización del lugar de trabajo o las relaciones externas a la misma. En la presente investigación el análisis está centrado en los diversos factores asociados a la obtención de resultados innovadores en productos y/o procesos en firmas industriales de la provincia de Córdoba.

Existe gran interés en comprender las variables que afectan al proceso de innovación, ya que es ampliamente reconocido como uno de los factores clave de la competitividad y el crecimiento económico de largo plazo. La innovación es un elemento fundamental en el acceso a mercados de productos de mayor diferenciación, más dinámicos y de mayores precios. La misma, ayuda a eludir la dependencia del comercio exterior basada en productos básicos. En este sentido Arza, *et al.* (2017) encuentran evidencia empírica favorable a esta hipótesis en el caso argentino. La innovación propicia aumentos en la productividad, a través de los cuales, las firmas innovativas tienen mejores posibilidades de avanzar en su inserción externa. Adicionalmente, la incorporación de conocimiento a la producción es referenciada como uno de los caminos para poder lograr un incremento sistemático de salarios sin afectar los niveles de competitividad (Lugones *et al.*, 2007).

A nivel de empresa, la innovación puede traducirse en un uso más eficiente de los recursos, permitiendo generar mejoras genuinas y sostenibles en el tiempo. Las firmas que basan su competitividad en la búsqueda e incorporación de nuevas tecnologías pueden ver su impacto en una reducción de costos posibilitando una mayor competencia vía precios o

bien posicionándose en nichos o segmentos de mayor valorización con la consecuente mejora de rentabilidad. De esta manera, las mayores tasas de rentabilidad asociadas a la innovación ubican a las empresas en mejores condiciones relativas para realizar nuevos gastos en I+D y otras actividades de innovación y les permite así incrementar sus probabilidades de obtener nuevas innovaciones y reiniciar el ciclo (Barletta *et al.*, 2013).

Tal como destacan Heijs y Buesa (2016), la innovación ya no se considera como un proceso que implica un número reducido de personas, sino que se la ve como un proceso que necesariamente debe integrar el aporte de diversos actores, en el cual tienen fundamental importancia las interacciones y el ambiente externo en que se está inmerso.

Comprender cómo se desarrolla el fenómeno de innovación posibilitaría conocer las sendas de desarrollo que éste induce en el país, la región o las empresas. Y esto a su vez, es base fundamental para el diseño, implementación y evaluación de las políticas de apoyo, las cuales permiten, de ser necesario, modificar esta senda. Identificar las características y acciones a nivel de firma que impactan sobre la probabilidad de obtener resultados innovadores, permite el diseño y adecuación, tanto de las decisiones propias de las empresas, como de las medidas provenientes del sector público para la promoción de la innovación.

El presente trabajo, en base a la reciente información disponible con representatividad provincial, busca generar un aporte a la comprensión del proceso de innovación en las empresas industriales manufactureras de la provincia de Córdoba para el período 2011-2016. Los datos utilizados provienen de la “Encuesta de Innovación Tecnológica” (EIT), llevada a cabo por la Dirección General de Estadística y Censos de la provincia. Se utilizan las observaciones y mediciones presentadas por cada empresa a través del tiempo, constituyendo lo que se denomina datos de panel (longitudinales).

1.2 Preguntas de investigación

El presente trabajo está guiado por las siguientes preguntas de investigación:

1. *¿Cuáles son las principales variables que influyen sobre la probabilidad de que una empresa industrial obtenga resultados positivos en materia de innovación?*

Para modelar la probabilidad de obtener resultados innovadores mediante la consideración de un conjunto de factores característicos de las empresas y de su entorno se aplica un modelo logístico mixto. Se considera que la empresa obtuvo resultados innovadores cuando declara que logró un producto y/o proceso nuevo o con mejoras significativas novedosos para el mercado. La elección de utilizar un modelo logístico mixto responde en primer lugar, a que la variable respuesta es binaria e indica el grupo al que pertenece la empresa (innova= 1 / no innova=0). En segundo lugar, a que los datos presentan una estructura de agrupamiento por tratarse de medidas repetidas en el tiempo para cada unidad (empresa), por lo que no es posible su tratamiento a través de modelos como la regresión logística. El supuesto de independencia no se cumple, e ignorar esta relación omitiendo la importancia de los efectos de grupo, puede tornar inválida muchas de las técnicas estadísticas tradicionales utilizadas para estudiar la relación entre los datos.

2. *En el marco de un modelo logístico mixto, ¿es posible, mejorar la predicción de la variable respuesta para unidades que no formaron parte del ajuste del modelo?*

En el modelo logístico tradicional, compuesto solo por la parte fija, es simple obtener predicciones de la variable respuesta a través de la función de enlace. En el caso de los modelos multinivel o mixto, la predicción de la variable respuesta en nuevas unidades se complejiza. Además del efecto fijo, el modelo cuenta con los efectos aleatorios estimados para cada grupo de la muestra con la que se ajustó el modelo (empresas en nuestro caso).

Avanzando en este sentido, se utiliza la parte fija de modelo en conjunto con una predicción de los efectos aleatorios realizada mediante diferentes métodos, para clasificar a nuevas empresas en innovadoras o no. Luego, se comparan los resultados obtenidos con los distintos métodos de estimación de los efectos aleatorios.

1.3 Objetivos

Esta investigación persigue dos objetivos generales. En primer lugar, modelar a nivel de firma los principales factores determinantes de la innovación en productos y/o procesos. En segundo lugar, en el marco de un modelo logístico mixto, predecir la probabilidad de obtener resultados innovadores en empresas que no formaron parte de la muestra con la que se ajustó el modelo.

Objetivos Específicos

1. Identificar las principales características de las empresas manufactureras innovadoras en la provincia de Córdoba, para el periodo 2011-2016.
2. Evaluar los principales determinantes de la innovación considerados en la literatura, en el caso de las empresas bajo estudio.
3. Analizar y comparar diferentes métodos de predicción de los efectos aleatorios, tal que permitan obtener la probabilidad de que una empresa no incluida en el ajuste del modelo (nueva empresa) sea innovadora.
4. Elaborar conclusiones que orienten la toma de decisiones empresariales y que permitan el diseño y orientación de políticas públicas en lo referente a innovación.

1.4 Aportes del trabajo

El análisis empírico de los datos sobre innovación mejora nuestro conocimiento sobre la misma, la evaluación de las políticas y la formulación de nuevas iniciativas (OECD, 2005). En este sentido, con la presente investigación, se espera contribuir en la comprensión de este complejo proceso, orientando la toma de decisiones tanto en el ámbito empresarial como en el de formulación de políticas públicas.

El trabajo se desarrolla sobre la base de información proveniente de la encuesta de innovación tecnológica de la provincia de Córdoba, convirtiéndolo en el primer análisis sobre esta fuente de datos con representatividad sub - nacional.

Es de destacar que el estudio se realiza con datos de un período reciente (2011- 2016). A nivel nacional, el Instituto Nacional de Estadística y Censos (INDEC) ha publicado cuadros con los principales resultados de la encuesta nacional hasta el año 2010. En tanto, el último informe corresponde al análisis de resultados del año 2005. Por su parte la Secretaría de Ciencia, Tecnología e Innovación Productiva de la Nación realiza un relevamiento sobre un panel de empresas potenciales ejecutoras de I+D del sector agropecuario, manufacturero y de servicios, utilizando una metodología diferente a la utilizada por la oficina de estadística provincial y por el INDEC. Los últimos resultados publicados son referidos al período 2014-2016.

Adicionalmente, se utiliza un modelo con efectos aleatorios para evaluar los principales determinantes de la probabilidad de innovar. Esto permite incorporar al análisis las mediciones presentadas por cada empresa a través del tiempo, siendo éste un importante aporte de la presente investigación.

Por último, se avanza en la comparación de diferentes técnicas para la predicción de los efectos aleatorios, ya que constituyen una dificultad en la clasificación de nuevas unidades (empresas) que no formaron parte de la muestra de estimación del modelo.

El trabajo se encuentra dividido en 5 capítulos. A continuación de la presente introducción, en el capítulo 2, se expone una revisión de un conjunto de trabajos teóricos y empíricos sobre la relación entre las características de las empresas y su entorno con la innovación. En el capítulo 3 se describe la fuente de datos, las principales variables que se incorporan al análisis; y se presentan las técnicas estadísticas utilizadas en el trabajo de aplicación. En el capítulo 4 se realiza un análisis descriptivo de las empresas analizadas, se detallan los resultados obtenidos en el ajuste del modelo, y en la aplicación de los distintos métodos de estimación de los efectos aleatorios para nuevas empresas. Para finalizar, en el capítulo 5 se plantean las conclusiones obtenidas, las limitaciones del trabajo y futuras líneas de investigación.

Capítulo 2: Marco teórico

En el presente capítulo, se examinan los principales enfoques teóricos sobre los factores determinantes de la innovación empresarial, planteando el marco conceptual sobre el cual se sustentan las hipótesis evaluadas para explicar la conducta innovativa de las empresas bajo análisis. Las visiones presentadas a continuación enfatizan diversos aspectos del proceso de innovación y pueden ser consideradas complementarias entre sí. Si bien existe consenso sobre el reconocimiento de la innovación y el cambio tecnológico como fuente fundamental del desarrollo económico, la gran cantidad de aristas de este fenómeno ha llevado a que los factores determinantes de la innovación sean aún objeto de estudio.

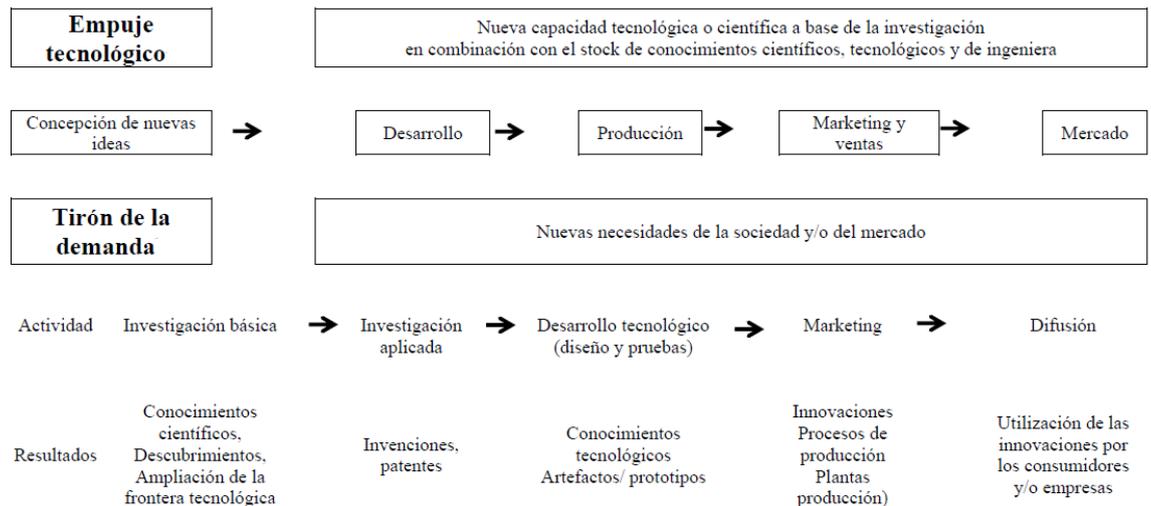
Los modelos de innovación fueron evolucionando en el tiempo, pasando de aquellos que explican la innovación como una relación lineal, hacia una visión evolucionista que entiende a la innovación como un proceso dinámico y complejo. Luego, a inicios de los años 90's se incorpora la perspectiva sistémica al análisis del proceso de innovación.

2.1 Modelo lineal

Hasta mitad de la década de los 70, predominaba el modelo explicativo de la innovación de carácter lineal. Tal como mencionan Heijs y Buesa (2016), la teoría económica consideraba que el proceso de producción de tecnología era resultado de la acción secuencial de las instituciones de investigación y de las empresas innovadoras. Basado en fundamentos neoclásicos, este modelo, concibe a la innovación como un proceso secuencial, unidireccional y ordenado que pasa por diferentes etapas, el cual inicia con la fase de investigación básica y finaliza con la fase de introducción de las innovaciones en el mercado.

Desde esta visión de la innovación, la literatura confrontó el rol de las oportunidades tecnológicas con el de la demanda. Estos modelos son también conocidos como de “primera y segunda generación”, respectivamente (Rothwell, 1994).

Gráfico 2.1: Esquema descriptivo del modelo lineal



Fuente: Heijs y Buesa (2016)

La primera posición conocida como “empuje tecnológico” (*technology-push*), contempla el desarrollo del proceso de innovación mediante un proceso secuencial y ordenado en el cual las nuevas tecnologías y la innovación son motorizadas por los descubrimientos científicos provenientes de la investigación básica. A partir del conocimiento científico (fuente de la innovación), pasando por la investigación aplicada, el desarrollo tecnológico, la fabricación y el lanzamiento al mercado de la novedad, se comercializa un producto o proceso que puede ser económicamente viable. Se plantea que las actividades de innovación son estimuladas principalmente por los avances científicos y tecnológicos (Rosenberg, 1974).

La segunda posición intensifica la atención puesta en el papel desempeñado por el mercado en el proceso innovador, lo que conduce a la emergencia de otro modelo (Rothwell, 1994). La hipótesis del “tirón de la demanda” (*demand-pull*) sostiene que existe

un vínculo positivo entre la actividad de innovación y el crecimiento del mercado (Schmookler, 1962). Bajo esta perspectiva, la innovación deriva principalmente del análisis de las necesidades de los consumidores. Éstas se convierten en la principal fuente de ideas para desencadenar el proceso de innovación. Las empresas pueden tomar el riesgo de desarrollar productos y procesos nuevos o mejorados si esperan niveles crecientes de demanda, o si clientes más sofisticados lo solicitan.

Heijs y Buesa (2016), destacan que estos modelos suponen que la transferencia tecnológica consiste en un proceso automático sin costos significativos o retrasos en el tiempo, basado en el mecanismo de “la mano invisible”. Las políticas públicas basadas en el modelo lineal estaban dirigidas fundamentalmente hacia la generación de innovaciones mediante la creación de centros de investigación, el apoyo a la I+D básica para tecnologías claves, o la financiación directa de las actividades de investigación empresariales.

Aunque simplistas en sus consideraciones, estas visiones establecieron las bases de los modelos posteriores. Algunas deficiencias de los modelos lineales, remarcadas por Velasco, *et al.* (2007), es que explican el proceso de innovación sin atender la interacción, retroalimentación y solapamiento de las distintas etapas del proceso. Por su parte, Heijs y Buesa (2016) destacan que no tienen en cuenta factores como el rol de las instituciones, de las estrategias y actitudes competitivas de otras empresas o países, o los factores relacionados a la educación.

2.2 Modelo evolucionista

El modelo evolucionista, a diferencia del modelo lineal, está basado en una visión sistémica y holística del proceso de innovación. Mientras el modelo lineal se focaliza en las actividades tecnológicas del departamento de I+D, el modelo interactivo o evolucionista destaca las capacidades tecnológicas de la empresa en general, considerando la gestión

de la innovación como un proceso estratégico y corporativo donde tendría que estar implicada toda la empresa, incluidos sus distribuidores y clientes (Heijs y Buesa, 2016).

Nelson y Winter (1982) plantearon un modelo evolutivo de innovación tecnológica gobernado por mecanismos de variación y selección. Dosi y Nelson (1994) señalan que el enfoque de la economía evolucionista estudia trayectorias temporales a partir de tres componentes fundamentales: i) las unidades de selección, ii) los mecanismos y criterios de dicha selección, iii) la adaptación y variación que deriva. En ese sentido Nelson (1995) explica que el enfoque evolutivo busca en las ciencias naturales un conjunto de similitudes y semejanzas válidas para explicar los procesos de transformación y desarrollo económico.

La literatura sobre innovación reconoce en el trabajo de Schumpeter un punto de referencia a partir del cual se desarrollaron las ideas evolucionistas. En este sentido Gutiérrez Rojas y Baumert (2018) señalan que este autor allanó el camino para el desarrollo del enfoque evolucionista de la innovación superando la teoría neoclásica, poniendo de relieve la importancia de la destrucción y acumulación creativa y resaltando el papel del empresario.

Schumpeter (1939) realizó una profunda crítica a la economía neoclásica y su definición de innovación, entendiendo a ésta última como causa motora del desarrollo económico. Considera que el desarrollo económico está movido por la innovación, por medio de un proceso dinámico de “destrucción creativa” en el cual nuevas tecnologías sustituyen a las antiguas. En este marco se considera la innovación como un fenómeno con múltiples expresiones impulsada por agentes concretos: los “emprendedores”. Algunos de los elementos presentes en el planteo schumpeteriano, como el tamaño de las empresas y el grado de concentración del mercado, son tomados como variables tradicionales de análisis sobre los determinantes de la innovación. Schumpeter (1942), plantea que la innovación es favorecida por el oligopolio. Las grandes firmas tienen mayores incentivos a innovar dada su posición de mercado y a la capacidad para generar recursos extra para destinarlos a las actividades de I+D. En este sentido, tal como destaca Gutiérrez Rojas y Baumert (2018), Schumpeter reconoció la importancia para el desarrollo innovador, tanto del nivel

de concentración del mercado como del rol de las pequeñas empresas y emprendedores individuales.

El planteo schumpeteriano ha sido heredado en sus lineamientos esenciales por la moderna corriente evolucionista del pensamiento económico cuyos principales desarrollos se refieren al análisis de los procesos de cambio tecnológico (Heijs y Buesa, 2016). El evolucionismo intenta explicar, desde una visión dinámica y endógena, los procesos y la evolución que las empresas tienen en el mercado. La innovación tecnológica es vista como un proceso evolutivo en el que se desarrolla una creciente diversidad de opciones a través de la creación por mutación de nuevas oportunidades, y la selección dentro del nuevo conjunto de opciones disponibles. Las opciones existentes pueden ser cambiadas por soluciones mejoradas o por la creación de opciones radicalmente nuevas que pueden coexistir con las reemplazadas (Dosi, 1997).

En este enfoque se encuentran presentes elementos como el aprendizaje, el conocimiento y las competencias poniendo un fuerte énfasis en la dinámica y en los procesos. La tecnología que utiliza la firma en la actualidad depende de la tecnología usada en el pasado. Juega un papel central en el proceso innovador el aprendizaje, donde las empresas no solo se nutren de la acumulación de experiencia, sino también de fuentes externas como consumidores, universidades, consultores, centros de investigación y competidores (Lundvall, 1992). En este contexto la innovación constituye un proceso de acumulación de conocimiento a través de las actividades propias y de la interacción con el ambiente en el que actúan los agentes. El aprendizaje se refiere a la creación de nuevas competencias y al establecimiento de nuevas habilidades y no a la obtención de la información (Lundvall y Johnson, 1994).

Otro factor importante es la capacidad de absorción de las firmas, es decir la habilidad para reconocer el valor de una nueva información, asimilarla y aplicarla con fines comerciales. La misma es función de los conocimientos tecnológicos previos, de la inversión en investigación y desarrollo, del aprendizaje en el proceso de fabricación y diseño, y de la

capacitación de los recursos humanos (Cohen y Levinthal, 1989). Esta habilidad está determinada por un amplio número de factores como la influencia de personas clave en el proceso de innovación, la cultura innovadora de la empresa, la capacidad y el proceso de aprendizaje mediante la acumulación de experiencias internas en la empresa, la estructura organizativa, y la gestión de la innovación, la integración del proceso de innovación en la estrategia general de la empresa, y las relaciones con el entorno (Heijs y Buesa 2016).

En el comportamiento innovativo de la firma, otro determinante relevante es el grado de apropiabilidad de los resultados de innovación. Éste consiste en las posibilidades de proteger las innovaciones de la imitación y de cosechar los beneficios de las actividades innovadoras (Levin *et al.*, 1987). El desarrollo tecnológico involucra diversos grados de especificidad, complementariedad e independencia que puede variar de manera importante entre sectores. Pavitt, *et al.* (1989) identifican patrones sectoriales en base a las oportunidades tecnológicas y niveles de apropiabilidad.

Como destaca Malerba (2002), la teoría evolutiva pone el énfasis en la dinámica, los procesos y las transformaciones. El aprendizaje y el conocimiento son elementos clave en el cambio de los sistemas económicos. Los agentes actúan, aprenden, investigan bajo incertidumbre en un ambiente cambiante. Adicionalmente, la teoría evolutiva pone énfasis en las creencias, objetivos y expectativas, a su vez afectados por el aprendizaje previo, la experiencia y por el entorno en el que actúan los agentes (Nelson, 1995; Dosi, 1997).

2.3 Sistemas de innovación

Partiendo de las ideas evolucionistas, a inicios de los años 90's, se fue abordando el proceso innovador desde una perspectiva sistémica (Freeman, 1987; Lundvall, 1992; Nelson, 1993; Edquist, 1997; entre otros). Desde esta visión toma relevancia la influencia de las instituciones externas, definidas en sentido amplio, sobre las actividades innovadoras de las empresas y de los otros agentes. Se acentúa la importancia de la

transferencia y difusión de las ideas, el conocimiento y la experiencia. En el enfoque de los sistemas de innovación, las interrelaciones y las redes de cooperación son los elementos fundamentales en el proceso de innovación y producción (Edquist, 1997).

Se ha denominado sistema de innovación al conjunto de agentes, instituciones y normas, en el que se apoyan los procesos de incorporación de tecnología, el cual determina el ritmo de generación, adaptación, adquisición y difusión de conocimientos tecnológicos en todas las actividades productivas (Nelson, 1993). Desde esta perspectiva, se sostiene que las firmas no innovan de forma aislada, el proceso de innovación involucra tanto a empresas como un sistema de interacciones e interdependencias entre las firmas y otras organizaciones e instituciones. De esta forma la innovación se deriva de redes tecnológicas.

Este enfoque subraya la importancia que tienen las fuentes de información externas a la empresa: los clientes, proveedores, consultoras, institutos de investigación, agencias gubernamentales, universidades, instituciones financieras, etc. Según Freeman (1987) un sistema de innovación se define como “las redes e instituciones en el sector privado y público cuyas actividades e interacciones inician, transmiten, modifican y difunden nuevas tecnologías”.

Dependiendo del objetivo y nivel de análisis, los sistemas de innovación, pueden ser “nacionales”, donde el foco se encuentra dentro de la frontera nacional e incluye además de las empresas las diferentes instituciones y organizaciones (Freeman, 1987; Nelson, 1993; Lundvall, 1992); “sistemas regionales/locales de innovación” donde el análisis se centra en la región (Cooke *et al.*, 1997); “sistemas tecnológicos de innovación” con el foco puesto en las redes de agentes para la generación, difusión, y utilización de las tecnologías (Hughes, 1984; Callon, 1992; Carlsson y Stankiewicz, 1995) y “sistemas sectoriales de innovación”, concentrado en las distintas industrias y complementario de los enfoques anteriores (Malerba, 2002).

Gutiérrez Rojas y Baumert, 2018, señalan que independientemente de si se está definiendo el “sistema nacional o regional de innovación”, ambos conceptos parten de un determinado enfoque geográfico, coincidiendo con ello de manera implícita en una relación destacada con el entorno económico, político y sociocultural.

Por su parte, el concepto de “sistema sectorial de innovación” y producción provee una visión multidimensional, integrada y dinámica de los sectores. Los agentes e individuos que componen un sistema sectorial de innovación interactúan a través de procesos de comunicación, intercambio, cooperación y competencia y estas interacciones son moldeadas por las instituciones. Un sistema sectorial experimenta cambios y transformaciones a través de la evolución conjunta de sus diversos elementos (Malerba, 2002).

Más allá de las definiciones y enfoques de los sistemas de innovación, como puede observarse en la tabla 2.1, los diversos autores han señalado un conjunto de elementos fundamentales y determinantes en la generación de los sistemas de innovación.

Tabla 2.1: Elementos destacados de los sistemas de innovación

Autor	Elementos fundamentales
Lundvall, 1992	Organización empresarial, relaciones entre empresas, gastos en I+D del sector público, estructura del sector financiero, organización e intensidad del I+D empresarial.
Nelson, 1993	Estructuras nacionales, incentivos a la innovación, capacidad creativa de los agentes económicos, y las singularidades culturales.
Patel y Pavitt, 1994	Empresas, centros de educación superior e investigación aplicada, y la administración pública.
Edquist y Jhonson, 1997	Sector privado (empresas), y sector público en su conjunto.

Fuente: Elaboración propia en base a Gutiérrez Rojas y Baumert (2018)

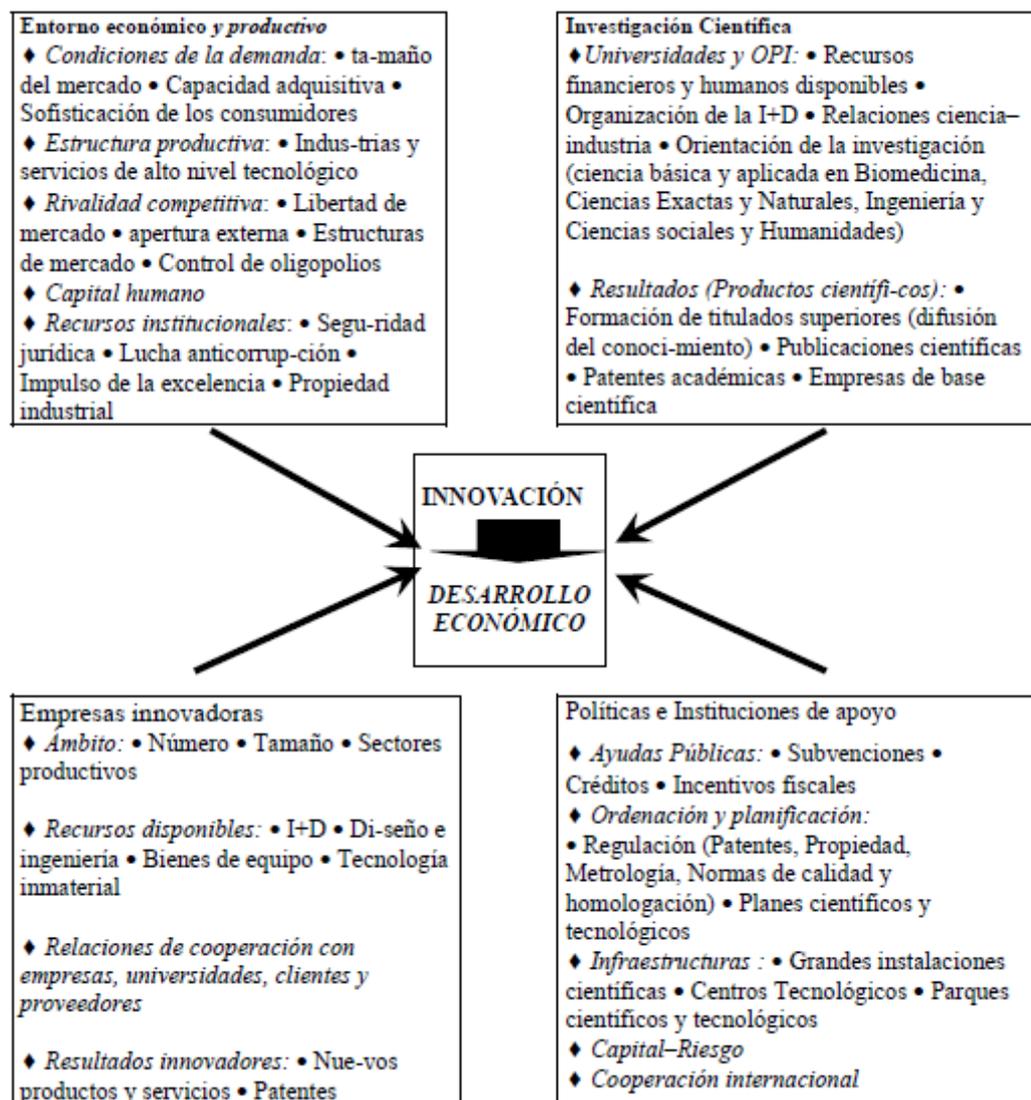
Por su parte, Gutiérrez Rojas y Baumert (2018) señalan como elementos principales de los sistemas de innovación tecnológica a las empresas con sus relaciones interempresariales y las estructuras de mercado, las actuaciones públicas en relación con la innovación y el

desarrollo tecnológico, la infraestructura pública y privada de soporte a la innovación, el entorno nacional y regional.

En línea con lo anterior, Heijs y Buesa (2016), proponen estructurar el análisis de los sistemas de innovación en torno a cuatro elementos principales (gráfico 2.2). En primer lugar, el que hace referencia al entorno económico y productivo en el que se encuentran las organizaciones del sistema, cuyas variables fundamentales son las que se refieren al tamaño del mercado y al nivel de desarrollo relativo del país o región, así como a la estructura de las actividades de producción. En segundo término, el que se focaliza en las actividades de investigación científica realizadas por los organismos públicos de investigación y las universidades, donde sus actividades posibilitan el progreso del conocimiento abstracto y, partiendo de él, la mejora de la formación y cualificación de la fuerza de trabajo, así como la resolución de los problemas concretos asociados a la producción. En tercer lugar, el que incluye a las empresas innovadoras y a su papel en el desarrollo tecnológico, tanto desde la perspectiva de la creación de las tecnologías que se emplean en los procesos de producción, como desde la de su adopción y difusión. Y, finalmente, el que concierne a las políticas que corrigen los fallos de mercado que afectan a la asignación de recursos a las actividades de creación de conocimiento, y a las que sostienen las instituciones que favorecen la apropiación de los resultados de la innovación o que contribuyen a la interrelación y cooperación entre las organizaciones y agentes del sistema de innovación.

De esta forma los autores definen al sistema nacional de innovación como el conjunto de organizaciones de naturaleza institucional y empresarial que, dentro del territorio, interactúan entre sí con el objeto de asignar recursos a la realización de actividades orientadas a la generación y difusión de los conocimientos sobre los que se soportan las innovaciones – principalmente tecnológicas - que constituyen, en el sentido schumpeteriano, el fundamento del desarrollo económico (Heijs y Buesa, 2016)

Gráfico 2.2: Esquema descriptivo de los Sistemas Nacionales (Regionales) de Innovación



Fuente: Heijs y Buesa (2016)

Los sistemas de innovación, cuyo enfoque es producto de la escuela evolucionista, entiende este fenómeno con una perspectiva global de la economía y la sociedad. Bajo esta visión, se considera la innovación como un proceso interactivo entre una amplia variedad de factores, dando lugar a aportes de naturaleza interdisciplinar, destacando la importancia de las instituciones y de las regulaciones, y enfatizando las interacciones que se producen entre todos sus elementos.

2.4 Estudios empíricos recientes

La mayor disponibilidad de encuestas específicas sobre innovación ha permitido el desarrollo de numerosos estudios empíricos que analizan distintos aspectos del proceso de innovación a nivel de empresa. Mairesse y Mohnen (2010), presentan una revisión de los principales contenidos de los relevamientos de innovación y de la utilización de esos datos. Organizan la revisión de los estudios econométricos considerando tres grandes categorías, a menudo solapadas. Distinguen entre trabajos centrados en los determinantes de la innovación, estudios focalizados en los efectos de la innovación, y estudios sobre temas específicos como las complementariedades y dinámica de la innovación.

Entre las investigaciones que analizan el impacto de la innovación sobre alguna medida de desempeño económico como puede ser, la productividad, el empleo o la dinámica exportadora, se destaca el trabajo de Crépon, Duguet y Mairesse (1998). Utilizando datos para Francia, proponen un modelo (llamado CDM) compuesto por tres ecuaciones que relacionan los esfuerzos innovadores (*input* innovador), la concreción de estas actividades (*output* innovador) y la productividad de la firma. Utilizando este desarrollo como referencia, numerosos trabajos abordan la relación entre innovación y productividad, tanto para países desarrollados como en desarrollo¹. Lööf *et al.* (2001), ajustan el modelo CDM para Finlandia, Noruega y Suecia; Griffith *et al.* (2006) trabajan con datos para Francia, Alemania, España y Reino Unido; Raffo *et al.* (2008) lo hacen para un conjunto de países europeos y latinoamericanos; Benavente (2005) estima el modelo para Chile, Crespi y Zúñiga (2010) lo hacen para seis países de Latinoamérica, entre ellos Argentina. Otros trabajos, basados en el modelo CDM, para nuestro país son el de Chudnovsky *et al.* (2004), Arza y López (2010), Gómez y Borrastero (2018), entre otros.

¹ Se puede consultar Raffo *et al.* (2008) para una revisión de trabajos basados en el modelo CDM aplicados a países en desarrollo entre el 2005 y el 2007

Una revisión de trabajos que abordan el fenómeno de la innovación en el sector industrial desde diversos objetivos para el caso argentino, publicados entre 2002 y 2006, se puede consultar en Anlló, Lugones y Peirano, (2007). Trabajos recientes como el de Milesi *et al.* (2011) examinan los determinantes en la probabilidad y la intensidad del gasto en I+D y en la adquisición de maquinarias y equipo. Por su parte, el trabajo de Marín *et al.* (2017) estudia la existencia de diferencias regionales en las oportunidades para la innovación en Argentina, utilizando un modelo logístico mixto, controlando por sector y región. El trabajo de Arza *et al.* (2017) aborda la relación entre las estrategias de innovación tecnológica adoptada por las empresas y sus capacidades exportadoras. Petelski *et al.* (2017) analizan el impacto del financiamiento público sobre los esfuerzos en innovación de las pymes argentinas. La investigación realizada por Astudillo (2018), se enfoca en un análisis comparativo de las características del proceso de innovación en las micro, pequeñas y medianas empresas manufactureras de Argentina y Ecuador.

Tal como se desprende de los párrafos anteriores, la innovación es un fenómeno de gran complejidad que puede ser abordado desde diferentes ángulos. En el presente trabajo nos concentramos en el estudio de los factores que influyen en la probabilidad de obtener resultados innovadores en productos y/o procesos, por parte de las empresas industriales de la provincia de Córdoba. Entre los determinantes de la innovación a nivel de firma, más allá de los objetivos buscados en los distintos estudios, la literatura pone énfasis en las características estructurales de las empresas, su comportamiento en materia de innovación y su entorno. Milesi *et al.* (2011), realizan una revisión de 29 trabajos empíricos en países desarrollados y en desarrollo, publicados entre 1997 y 2007, identificando un conjunto de determinantes de la innovación considerados en los mismos.

Uno de los factores ampliamente analizados es el tamaño de las empresas. En este sentido, existen trabajos que apoyan la conocida “hipótesis shumpeteriana” que considera a las firmas grandes, y con poder de mercado como las que mayores posibilidades innovadoras tienen. Tal como destacan Crespi y Zúñiga (2010), las economías de escala

relacionadas a la producción y a la diversificación de I+D, el mayor acceso al financiamiento y una mejor apropiación de la difusión de conocimientos externos, permiten a las empresas más grandes un mejor posicionamiento para aprovechar inversiones en innovación. Por su parte, Arias Ortiz *et al.* (2013), menciona otras investigaciones que refutan estas conclusiones, considerando que las empresas pequeñas son más flexibles, tienen mayor capacidad de adaptación y estructuras organizacionales menos complejas, lo que favorece a la innovación y al desarrollo de nuevos proyectos.

En la revisión bibliográfica realizada por Milesi *et al.* (2011) se destaca que la proporción de modelos en los que esta variable resultó tener efectos positivos sobre la probabilidad de la innovación es superior al 60% para las aplicaciones en países desarrollados, y del 85% en las aplicaciones de países en desarrollo. En diversos trabajos para Argentina, esta variable alterna su signo. Investigaciones como las de Chudnovsky *et al.* (2004), Crespi y Zúñiga (2010), encuentran que las empresas de mayor tamaño son más innovadoras. En tanto, trabajos como los de Bachmann (2016), Marin *et al.* (2017), Astudillo (2018), Gómez y Borrastero (2018), refutan estas conclusiones, encontrando relación negativa o no significativa entre tamaño y la obtención de resultados innovadores.

Otras variables incorporadas habitualmente en los estudios sobre determinantes de la innovación son aquellas relacionadas a los esfuerzos realizados por las empresas (*input* innovador), tanto en actividades de investigación y desarrollos (I+D) como en otras actividades de innovación (adquisición de maquinaria y equipo, hardware, software, contratación de tecnología, consultoría). Mairesse y Mohnen (2010) encuentran en su revisión de la literatura que la mayoría de los estudios para países desarrollados confirman la existencia de una relación positiva entre los gastos en investigación y desarrollo (I+D) y la obtención de resultados innovadores, en especial si son realizados de forma continua. Un ejemplo de ello es el estudio de Griffith *et al.*, (2006), para Francia, Alemania, España y Reino Unido. Resultados equivalentes se encuentran para el caso de economías en desarrollo. Hay evidencia de que mayores gastos en I+D conducen a una mayor propensión

a introducir innovaciones tecnológicas en empresas de Chile (Benavente, 2005), Brasil y México (Raffo *et al.*, 2008), Argentina (Chudnovsky *et al.*, 2004, Raffo *et al.*, 2008, Arza y López, 2010).

Crespi y Zúñiga, (2010) destacan que, en muchas economías latinoamericanas las actividades de innovación de las empresas están basadas en imitación y transferencia de tecnología, adquisición de maquinaria y equipos y compra de tecnología desincorporada. En su estudio incorporan tanto las actividades de innovación intra-muros (I+D interna e ingeniería y diseño) como las restantes actividades de innovación, encontrando un efecto positivo de todas las actividades sobre la obtención de resultados innovadores, para todos los países analizados. A igual resultado arriban las investigaciones para Argentina de Marin *et al.* (2017) y Gómez y Borrastero (2018).

Se esperan mayores efectos positivos de la inversión realizada en I+D si la misma se hace de manera continua y no esporádica. Buesa *et al.* (2002) destacan que las empresas que gastan en I+D en forma sistemática, son las que más aprovechan los efectos acumulativos y de aprendizaje derivados de la continuidad de la investigación tecnológica. Lo mismo concluyen Chudnovsky *et al.* (2004), y Marín, *et al.* (2017) quienes señalan la importancia de que las firmas aprendan al innovar y que dicho proceso de aprendizaje debe ser continuo para ser efectivo.

La inversión en actividades de innovación requiere de un gran esfuerzo por parte de las empresas. Características de este tipo de inversiones, como la incertidumbre, la indivisibilidad y la naturaleza intangible de las mismas, hacen difícil el acceso al crédito. Es muy difícil la tarea de negociar un préstamo con activos poco seguros y celebrar contratos con activos de conocimiento (intangibles, por definición) como garantía (Arias Ortiz *et al.*, 2013). De la revisión de estudios realizada por Mairesse y Monhen, (2010), surge que la mayor parte de ellos concluye que el financiamiento público de la investigación y desarrollo conduce a mayores inversiones en actividades de innovación y a mayores resultados innovadores y no desplaza gasto privado por público.

En los países latinoamericanos el apoyo gubernamental al financiamiento de la I+D es esencial. Crespi y Zúñiga (2010), plantean que los altos costos y riesgos de la innovación y la dificultad de las empresas de esperar largos períodos de tiempo para ver resultados son los principales obstáculos, a la innovación, percibidos por las empresas de la región. En su investigación concluyen que existe relación positiva entre financiamiento público y el gasto en actividades de innovación en el caso de Chile, Colombia y Costa Rica, pero la relación es no significativa para el caso de Argentina. Por el contrario, Petelski *et al.* (2017) en su estudio focalizado en el financiamiento público a la innovación en Argentina, encuentra evidencias de un impacto positivo del mismo sobre la intensidad del esfuerzo en I+D de las empresas industriales.

A menudo las empresas realizan actividades de innovación en colaboración con clientes, proveedores, universidades, laboratorios, etc. La cooperación permite compartir conocimientos, beneficiarse de complementariedades, reducir riesgos y ahorrar costos. Mairesse y Monhen (2010) encuentran en su revisión de trabajos para países industrializados que la cooperación en actividades de innovación (I+D o diseño) está asociada a mayores niveles de inversión en estas actividades. A igual conclusión arriba Raffo *et al.* (2008) para un conjunto de países europeos cuando la cooperación es internacional. En tanto para los países latinoamericanos sus conclusiones son dispares. Crespi y Zúñiga (2010) encuentran evidencia parcial para un conjunto de países latinoamericanos. La relación es significativa en el caso de Colombia, Panamá y Uruguay, pero no lo es para Argentina, Chile y Costa Rica, resultado que puede estar mostrando un débil desarrollo de las redes de innovación en estos países. Como se menciona en ese trabajo, la colaboración podría ser un proceso costoso que requiere horizontes de tiempo más prolongados en las economías en desarrollo.

Otras variables incluidas en gran parte de los trabajos empíricos revisados son la antigüedad de la empresa, el origen del capital, su inserción internacional, la capacitación de los recursos humanos, el sector industrial al que pertenece, entre otras. En la tabla 2.2

se enumeran los principales trabajos empíricos seleccionados que constituyen la revisión de la literatura en materia de innovación, para el presente trabajo

Tabla 2.2: Estudios empíricos seleccionados

Autor/es	Año	País/es	Objetivo
Crepon, B.; Duguet, E. y Mairesse, J.	1998	Francia	Estudiar los enlaces entre productividad, innovación e investigación (modelo CDM).
Lööf, H.; Heshmati, A.; Asplund, R. y Naas, S.O.	2001	Finlandia, Noruega y Suecia	Analizar la relación entre innovación y productividad a nivel de firma.
Buesa, M.; Baumert, T.; Heijs, J. y Martínez, M.	2002	España	Determinar los factores determinantes de la innovación, enfocando el estudio desde la visión de sistemas regionales de innovación.
Chudnovsky, D.; Lopez, A. y Pupato, G.	2004	Argentina	Analizar los determinantes de los inputs y outputs innovadores, y su impacto sobre la productividad laboral de las firmas industriales.
Benavente Hormazábal, J. M.	2005	Chile	Estudiar los determinantes de las actividades de investigación y desarrollo (I+D) y su impacto sobre la innovación tecnológica y la productividad de firmas.
Griffith, R.; Huergo, E.; Mairesse J. y Peters, B.	2006	Francia, Alemania, España, Reino Unido	Investigar el nexo entre innovación y productividad a nivel de firma, realizando una comparación entre países.
Anlló, G.; Lugones, G. y Peirano, F.	2007	Argentina	Analizar el perfil innovativo argentino, tanto para el agro como para la industria, abarcando el período 1992-2004.
Raffo, J.; L'huillery, S. y Miotti, L.	2008	Francia, España, Suiza, Argentina, Brasil y México	Comparar el rol de la innovación y el desempeño económico entre los países bajo análisis, modelando la relación entre intensidad de I+D, innovación y productividad.
Arza, V., López A.	2010	Argentina	Estimar la relación entre innovación y productividad y la realización de actividades de innovación en Argentina, para el período 1998-2004
Crespi, G. y Zuñiga, P.	2010	Argentina, Chile, Colombia, Costa Rica, Colombia, Panamá y Uruguay	Examinar los determinantes de la innovación tecnológica y su impacto sobre la productividad laboral.
Mairesse, J. y Mohnen, P.	2010	Países desarrollados y en desarrollo	Realizar propuestas que permitan avanzar tanto en el diseño como en la implementación de las encuestas de innovación, y sobre su utilización para el análisis econométrico, a partir de una revisión de la literatura sobre la temática.
Milesi, D.; Petelski, N. y Verre, V.	2011	Argentina	Examinar los determinantes que influyen sobre la probabilidad e intensidad del gasto en I+D y en maquinaria y equipo.
Arias Ortiz, E.; Crespi, G.; Tacsir, E.; Vargas, F. y Zuñiga, P.	2013	Región de América Latina y el Caribe	Describir, mediante el análisis de un conjunto de indicadores de innovación, el comportamiento de las empresas manufactureras en la región de América Latina y el Caribe.
Bachmann, F.	2016	Argentina	Determinar los factores asociados al éxito en la innovación en productos y procesos en firmas industriales.
Arza, V., Español, P. y Herrera Bartis, G.	2017	Argentina	Analizar la vinculación entre las estrategias de innovación tecnológica adoptadas por las empresas industriales argentinas y sus capacidades exportadoras entre 1998 y 2005.
Marín, A.; Liseras, N.; Calá, C. y Graña, F.	2017	Argentina	Investigar si existen diferencias regionales en las oportunidades para la innovación en Argentina y qué características regionales explican estas diferencias.
Petelski, N.; Milesi, D. y Verre, V.	2017	Argentina	Evaluar cómo impacta la política de financiamiento público a la innovación sobre la intensidad de los esfuerzos tecnológicos de las Pymes.
Astudillo Durán, S.	2018	Argentina y Ecuador	Comparar las actividades de innovación del sector manufacturero de Ecuador con el de Argentina.
Gómez, M. C. y Borrastero, C.	2018	Argentina	Analizar el impacto de la innovación de las empresas en su heterogeneidad productiva.

Fuente: Elaboración propia.

En base a lo expuesto en este capítulo, se puede concluir que la evidencia empírica sobre determinantes de la innovación es amplia y no muestra resultados concluyentes. Se

entiende al proceso de innovación como un fenómeno complejo en el que prevalecen relaciones dinámicas, donde los factores relacionados a la misma son muchos, variados y cambiantes. En esta investigación, nos centramos en evaluar e identificar las principales variables vinculadas a las empresas manufactureras de la provincia de Córdoba, que ayuden a explicar la probabilidad de que las firmas sean innovadoras en productos y/o procesos. Se analizan características de las empresas, variables relacionadas a su comportamiento en materia de innovación y al entorno donde desarrollan sus actividades.

Capítulo 3: Datos y Metodología

En el presente capítulo se describe la fuente de datos y las principales variables que se incorporan al análisis, continuando luego con la presentación de la metodología utilizada para evaluar los factores determinantes de la innovación y las técnicas que permiten completar la predicción para nuevos grupos que no formaron parte de la muestra con la que se estimó el modelo.

3.1 Fuente de datos: características de la encuesta

Los datos utilizados provienen de la Encuesta de Innovación y Conducta Tecnológica (EIT), para los años 2011 a 2016, realizada por la Dirección General de Estadística y Censos de la provincia de Córdoba. La misma posee representatividad provincial para el sector industrial.

El formulario de relevamiento es similar al de la Encuesta Nacional sobre Innovación y Conducta Tecnológica realizada por el Instituto Nacional de Estadística y Censos (INDEC). Tanto la encuesta nacional, como la realizada a nivel provincial responden a los conceptos y recomendaciones contenidos en los Manuales Frascati (2003) y Oslo (2005) de la OECD y en el Manual de Bogotá (2001) de la RICYT, los cuales son una base ineludible para la orientación de los ejercicios de medición de las actividades innovativas. El respeto por las definiciones contenidas en los manuales mencionados permite la comparabilidad nacional e internacional de los indicadores que surgen de la encuesta.

El objetivo principal de la encuesta es brindar información, representativa para la provincia de Córdoba, para la elaboración de indicadores de innovación, investigación y desarrollo (I+D) y sobre el uso de las Tecnologías de la Información y de las Comunicaciones (TICs), generando una herramienta que permita orientar estrategias privadas y políticas públicas en la materia.

El diseño muestral, el trabajo de campo, la sistematización, el análisis de la información y publicación de los resultados son realizados por personal de la Dirección General de Estadística y Censos de la Provincia de Córdoba. Para el diseño y realización de los primeros ejercicios de medición se contó con el apoyo del equipo técnico de INDEC y del Ministerio de Ciencia y Tecnología provincial. Los datos suministrados por las empresas están amparados por secreto estadístico, por lo que pueden ser proporcionados y publicados exclusivamente en compilaciones de conjunto y ser utilizada solo con fines estadísticos.

El formulario está compuesto de cinco secciones donde se consulta sobre datos de la empresa y desempeño económico, actividades de innovación, innovaciones logradas, tecnologías de la información y las comunicaciones (TIC's) y medio ambiente. En el formulario se encuentran explicadas las definiciones conceptuales y aclaraciones relevantes para que la empresa pueda completarlo. Adicionalmente, existe un mecanismo de consultas vía correo electrónico y telefónico, mediante el cual las empresas reciben asesoramiento de parte del equipo técnico de relevamiento para subsanar dificultades que se puedan presentar durante la respuesta del cuestionario.

La muestra teórica de la encuesta está constituida por 540 empresas del sector industrial manufacturero, distribuidas en todo el territorio provincial. La tasa de respuesta promedio que se obtuvo en el período 2011-2016 fue del 85%. El relevamiento de campo es realizado con periodicidad anual. La información de los años 2011 y 2012 fue obtenida mediante la distribución de cuestionarios a las empresas muestreadas y su posterior recuperación, alternativamente, a través de encuestadores y correo postal. A partir del año 2013 la encuesta se realiza únicamente en formato electrónico a través de un formulario que se envía por correo electrónico.

3.2 Muestra y variables

El total de empresas disponibles fue separado en dos muestras. Una para realizar el ajuste del modelo, compuesta por aquellas para las que se tenía información completa de los últimos cinco años analizados (muestra de entrenamiento) y otra para realizar inferencia sobre nuevos datos (muestra de validación), en la cual fueron incluidas la totalidad de las empresas restantes. En ambos grupos se incluyeron aquellas firmas que en el período analizado tuvieron, en promedio, entre 5 y 250 empleados (PyMes).

Los datos utilizados son las observaciones y mediciones presentadas por cada empresa a través del tiempo. En la muestra de entrenamiento se consideró información de todos los años disponibles y en el segundo grupo se utilizaron los valores del último año en que la empresa respondió la encuesta.

Tabla 3.1: Empresas y observaciones en la muestra

Muestra	Cantidad de empresas	Cantidad de observaciones
<i>Entrenamiento</i>	276	1.631
<i>Validación</i>	232	232
Total	508	

Para evaluar las principales variables consideradas relevantes en el marco teórico, se ajusta un modelo logístico mixto donde la variable respuesta considera si la empresa logró nuevos o mejoras significativas en productos y/o procesos novedosos para el mercado nacional y/o internacional. Esta variable es binaria e indica si la empresa innovó o no lo hizo (innova= 1 / no innova=0). En tanto, las variables predictoras, son indicadores vinculados a las características y comportamiento de las firmas, y a su relación con el sistema nacional de innovación. A continuación, se expone en la Tabla 3.2 las variables consideradas en el modelo con sus respectivas definiciones conceptuales.

Tabla 3.2: Variables incluidas en el modelo

Variable	Nombre	Descripción
Variable dependiente:		
<i>Innovación</i>	<i>INNOVA</i>	1, si la empresa logró productos y/o procesos nuevos o con mejoras significativas, novedosos para el mercado nacional y/o internacional. 0, en caso contrario
Variables predictoras:		
<i>Tamaño</i>	<i>LN_RRHH</i>	Tamaño de la empresa, medido con el personal ocupado (en logaritmo).
<i>Esfuerzos en actividades de innovación</i>	<i>AI_ENDO_VTA</i>	Porcentaje del gasto en actividades de innovación endógenas (investigación y desarrollo interna, actividades de capacitación y actividades de diseño industrial e ingeniería), en relación a las ventas totales.
	<i>AI_EXO_VTA</i>	Porcentaje del gasto en actividades de innovación exógenas (investigación y desarrollo externa, adquisición de maquinaria y equipo, hardware, software, contratación de tecnología, consultorías), en relación a las ventas totales.
	<i>AI_ENDO_CONT4</i>	1, si la empresa realizó gastos en I+D interna, capacitación y/o actividades de diseño industrial e ingeniería de manera continua durante 4 años o más. 0, caso contrario.
<i>Relación con el sistema nacional de innovación</i>	<i>FONDOS_PROGOFIC</i>	1, si la empresa utilizó fondos de programas oficiales de promoción de la innovación para financiar sus actividades. 0, caso contrario.
	<i>VINCULA</i>	1, si la empresa realizó alguna actividad de vinculación con entidades públicas o privadas del sistema de innovación nacional (Universidades, INTI, INTA, CONICET, otros organismos gubernamentales, clientes, proveedores, laboratorios, etc). 0, caso contrario.

En cuanto a la variable dependiente, el éxito en el proceso innovador es definido por la firma. En el cuestionario se solicita que la empresa declare si las actividades de innovación realizadas han conducido a logros innovadores en el año de referencia. Se identifican las innovaciones desagregando el objeto sobre el que se innovó (producto, procesos, organización, comercialización) y si fue novedoso para la firma, el mercado nacional o el internacional.

En la presente investigación se consideraron innovadoras a las empresas que declararon haber logrado productos y/o procesos nuevos o significativamente mejorados, novedosos

para el mercado nacional o internacional. El análisis se basa en un indicador subjetivo, ya que es autodeclarado por las empresas.

Se define a la innovación en *productos nuevos o mejorados* como: “la introducción al mercado de un producto tecnológicamente nuevo (cuyas especificaciones técnicas, componentes, materiales o características funcionales difieren significativamente de los correspondientes a productos anteriores de la empresa) o significativamente mejorado (previamente existente cuyo desempeño ha sido perfeccionado o mejorado en gran medida)”. Por otra parte, la innovación en *procesos nuevos o mejorados* implica recrear o modificar el proceso de elaboración de productos o la prestación de servicios, como resultado de utilizar nuevos equipos, insumos, soluciones tecnológicas o de introducir cambios en la organización del proceso productivo. Incluye modificaciones en la logística de insumos o de productos terminados (OECD, 2005).

3.3. Método

Para identificar y evaluar factores relevantes en el proceso de innovación en las firmas industriales de la provincia se utiliza un modelo de regresión logística con efectos aleatorios. Este modelo, perteneciente al grupo de modelos lineales generalizados mixtos, permite trabajar con variable respuesta dicotómica y observaciones correlacionadas entre sí. A continuación, se detallan los rasgos más relevantes del modelo utilizado para evaluar los determinantes de la innovación; luego se describen las técnicas que permiten completar la predicción de la variable respuesta en los modelos mixtos y finalmente se presentan medidas utilizadas para evaluar el desempeño del modelo y comparar los métodos de estimación de los efectos aleatorios.

3.3.1 Modelos para respuesta no Normal

Siguiendo lo expuesto por McCullagh y Nelder (1989), los modelos lineales generalizados (MLG) proveen una extensión de los modelos lineales clásicos en dos sentidos. En primer lugar, relaja el supuesto de normalidad de la variable dependiente, permitiendo que la distribución del componente aleatorio del modelo sea cualquier función perteneciente a la familia exponencial, donde la distribución Normal es solo un caso. En segundo lugar, relaja el supuesto de relación lineal entre los efectos de las covariables y la media de la variable respuesta.

Un MLG puede ser especificado por tres componentes:

- a) Un componente aleatorio, representado por las variables aleatorias independientes y_i , provenientes de la misma distribución en la familia exponencial, en su forma canónica dada por:

$$f(y_i, \theta_i, \phi) = \exp\{[y_i \theta_i - b(\theta_i)]/a(\phi) + c(y_i; \phi)\} \quad (3.1)$$

Para $a(\cdot)$, $b(\cdot)$, y $c(\cdot)$ funciones conocidas. El parámetro θ_i es llamado parámetro canónico, o natural de la familia exponencial, y ϕ es llamado parámetro de dispersión.

Donde:

$$E(Y_i) = \mu_i = b'(\theta_i)$$

$$V(Y_i) = a(\phi)b''(\theta_i) = a(\phi)V(\mu_i)$$

$$V(\mu_i) = \frac{d\mu_i}{d\theta} \text{ es la función de varianza}$$

- b) Un componente sistemático, lineal en los parámetros, llamado predictor lineal, que especifica la combinación lineal de las variables explicativas,

$$\eta_i = x_i\beta \quad (3.2)$$

- c) Una función de enlace $g(\cdot)$, monótona y diferenciable, que relaciona el componente sistemático y el valor esperado del componente aleatorio, tal que

$$\eta_i = g(\mu_i) \quad (3.3)$$

Cada función de distribución de la variable respuesta permite una clase de funciones de enlace. La función de enlace identidad, tal que $E(Y) = \mu = \eta$, es apropiada siempre y cuando μ y η puedan tomar cualquier valor en \mathfrak{R} . Esto, por ejemplo, no sucede en el caso de datos de conteo con distribución Poisson donde $\mu \geq 0$, o en datos con distribución Binomial donde $0 \leq \mu \leq 1$.

Cada distribución de la familia exponencial posee una función de enlace especial para la cual existe estadístico suficiente con igual dimensión que el vector de parámetros β en el predictor lineal. Éstas funciones de enlace se denominan canónicas, y son aquellas tal que el parámetro natural coincide con el predictor lineal: $\theta_i = \eta_i$.

Alguna de las funciones de distribución utilizadas habitualmente, pertenecientes a la familia exponencial, con sus respectivas funciones de enlace canónicas, son: Normal con función de enlace Identidad, distribución Poisson con función de enlace Log, distribución Gamma con función enlace Inversa y distribución Binomial con función de enlace Logit.

Otras funciones de enlace, distinta a la canónica, que pueden ser utilizadas para modelos con respuesta binaria son probit y complemento log-log. La función de enlace probit genera estimaciones similares a la logit y es definida como la inversa de la distribución normal estándar. La función de enlace complemento log-log está indicada para distribuciones con valores extremos o distribuciones asimétricas en los datos.

A pesar de que sea conveniente utilizar la función de enlace canónica, ya que la misma conduce a propiedades estadísticas del modelo deseables, en particular cuando las muestras son pequeñas, no hay razón a priori por la cual el componente sistemático deba ser aditivo en la escala dada por esa función de enlace. La selección de la función de enlace se basa en la búsqueda de simplicidad y habilidad para interpretar los resultados de análisis y en el ajuste apropiado del modelo (Skcrondal y Rabe-Hesketh, 2004).

3.3.1.1 Modelo logístico

Uno de los objetivos centrales del presente trabajo es investigar la relación entre la probabilidad de que las empresas logren resultados innovadores en productos y/o procesos y un conjunto de variables explicativas. En estas situaciones, donde la variable respuesta Y_i solo puede asumir dos valores posibles, éxito (1) o fracaso (0), es usual utilizar un modelo de regresión logística, el cual utiliza como función de enlace la función logit.

Sean y_i con $i = 1, \dots, q$, variables dicotómicas independientes, que siguen una distribución Bernoulli con parámetro μ_i , tal que $\Pr(y_i = 1) = \mu_i$, su función de densidad perteneciente a la familia exponencial dada en (3.1), es tal que:

$$a(\phi) = 1$$

$$b(\theta_i) = \ln(1 + e^{\theta_i})$$

$$\theta_i = \ln\left(\frac{\mu_i}{1 - \mu_i}\right) \text{ función de enlace logit}$$

La esperanza y la función de varianza quedan definidas por:

$$E(Y_i) = b'(\theta_i) = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$$

$$V(Y_i) = a(\phi)b''(\theta_i) = a(\phi)V(\mu_i) = \mu_i(1 - \mu_i)$$

El conjunto de variables explicativas $\mathbf{x}_i = (x_1, x_2, \dots, x_p)$ influyen sobre la variable respuesta a través del predictor lineal $\eta_i = \mathbf{x}_i\boldsymbol{\beta}$ y se conecta con el valor esperado del componente aleatorio mediante la función de enlace logit. Ésta se define como el logaritmo del cociente - $\ln(odds)$ - entre la probabilidad de que el evento ocurra (éxito) y la probabilidad de que no ocurra (fracaso):

$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{x}_i\boldsymbol{\beta}$$

Luego, utilizando la inversa de la función de enlace, la probabilidad de obtener una respuesta positiva es:

$$\Pr(y_i = 1 | \mathbf{x}_i) = \mu_i = \frac{\exp\{\mathbf{x}_i\boldsymbol{\beta}\}}{1 + \exp\{\mathbf{x}_i\boldsymbol{\beta}\}}$$

Los coeficientes estimados se interpretan como el cambio en el logaritmo de las chances producido por un cambio unitario en la variable independiente asociada, permaneciendo las demás variables constantes. Es decir, si la variable $x_1=a$, aumenta en una unidad $x_1=a+1$, luego:

$$\begin{aligned} \ln\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a+1} - \ln\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a} &= (\beta_0 + \beta_1 * (a + 1) + \dots + \beta_p x_p) - (\beta_0 + \beta_1 * (a) + \dots + \beta_p x_p) \\ &= \beta_1 \end{aligned}$$

De donde se obtiene que el cociente de chances (odds-ratio), generado por un incremento unitario en la variable x_1 , manteniendo las demás constantes, queda expresado como:

$$\begin{aligned} \exp\left\{\ln\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a+1} - \ln\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a}\right\} &= \frac{\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a+1}}{\left(\frac{\mu_i}{1-\mu_i}\right)\Big|_{x_1=a}} \\ &= \exp(\beta_1) \end{aligned}$$

3.3.1.2 Modelo logístico mixto

Uno de los supuestos del modelo de regresión logística es que las observaciones de la variable respuesta son independientes entre sí, pero existe un gran número de situaciones en las cuales los datos presentan una estructura de cluster o de niveles múltiples. Ejemplo de esto es cuando se recopila información de los individuos o unidades de análisis en distintos momentos del tiempo (medidas repetidas/datos longitudinales), o cuando los individuos están anidados en unidades mayores, integran familias, grupos, clases. Este agrupamiento, cualquiera sea su origen, provoca que las observaciones tiendan a estar correlacionadas dentro de los grupos ya que las unidades comparten un mismo ambiente o presentan características semejantes. El supuesto de independencia no se cumple, e ignorar esta relación, omitiendo la importancia de la dependencia intra-grupo, puede conducir a inferencias incorrectas.

Los datos analizados en el presente trabajo son las mediciones realizadas anualmente, en el período 2011- 2016, sobre una muestra de empresas industriales con actividad en la provincia de Córdoba. Debido a que la estructura de los datos presenta dependencia en las respuestas múltiples de cada empresa, no es posible su tratamiento a través de un modelo de regresión logística como el presentado en el apartado anterior. La incorporación de efectos aleatorios en la regresión permite modelar la heterogeneidad no observada a nivel de unidad (empresa).

Los modelos lineales generalizados mixtos (MLGM), hacen posible modelar situaciones donde la variable respuesta tiene una distribución no normal, incorporando efectos aleatorios para las unidades/grupos. En estos modelos, como las unidades de cada nivel son consideradas muestras aleatorias de una población, contribuyen en el modelo incorporando efectos aleatorios. Estos últimos, permiten coeficientes de regresión aleatorios que reflejan la variabilidad intra-unidades, mediante la variabilidad de las ordenadas al origen y/o de las pendientes.

Sea y_i el vector de respuestas del sujeto i , donde y_{ij} es la respuesta en el momento j del sujeto i , tal que $i = 1, \dots, q$, y $j = 1, \dots, t_i$. Se supone que, condicionado sobre los efectos aleatorios α_i , los elementos y_{ij} son independientes. Todas las y_{ij} tienen función de densidad de la forma:

$$f_{ij}(y_{ij} | \alpha_i, \theta_{ij}, \phi) = \exp\{[y_{ij} \theta_{ij} - b(\theta_{ij})]/a(\phi) + c(y_{ij}; \phi)\} \quad (3.4)$$

La media condicional $E(y_{ij} | \alpha_i) = \mu_{ij}$ es modelada a través del predictor lineal η_{ij} , que se forma como una combinación de efectos fijos y aleatorios,

$$\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\alpha_i \quad (3.5)$$

y una función de enlace conocida $g(\mu_{ij})$ que relaciona la media condicional con el predictor lineal. Los vectores \mathbf{x}_{ij} y \mathbf{z}_{ij} contienen los valores conocidos de las covariables, asociados con $\boldsymbol{\beta}$ ($p \times 1$) vector de efectos fijos y con α_i ($k \times 1$) vector de efectos aleatorios del cluster

i , respectivamente. Adicionalmente, se supone que α_i se distribuye normal multivariada con media 0 y matriz de covarianzas Σ , y denotamos su densidad como $f(\alpha_i | \Sigma)$.

El modelo puede ser ajustado maximizando la verosimilitud marginal, obtenida integrando sobre los efectos aleatorios. La contribución a la verosimilitud del sujeto i , resulta:

$$f_i(y_i | \beta, \Sigma) = \int \prod_{j=1}^{t_i} f_{ij}(y_{ij} | \alpha_i, \beta) f(\alpha_i | \Sigma) d\alpha_i \quad (3.6)$$

De donde se obtiene la verosimilitud para β y Σ como:

$$L(\beta, \Sigma) = \prod_{i=1}^q f_i(y_i | \beta, \Sigma) = \prod_{i=1}^q \int \prod_{j=1}^{t_i} f_{ij}(y_{ij} | \alpha_i, \beta) f(\alpha_i | \Sigma) d\alpha_i \quad (3.7)$$

Generalmente, las integrales en (3.7) no presentan forma cerrada. Para maximizar la verosimilitud es necesario realizar una aproximación a la integral en cada paso de optimización. Algunos de los métodos que pueden utilizarse son el de Laplace, cuadratura numérica de Gauss-Hermite y de cuadratura adaptativa. Cuando la estructura de los efectos aleatorios es compleja estos métodos no funcionan adecuadamente, para lo cual han sido propuestas una variedad de técnicas usando ideas de cuasi-verosimilitud (*QL*). Entre estos métodos podemos mencionar el de cuasi-verosimilitud penalizada (*PQL*) (Schall, 1991; Breslow y Clayton, 1993), verosimilitud – jerárquica (*h-likelihood*) (Lee y Nelder 1996, 2001). Puede consultarse una revisión de los diferentes métodos de estimación en Skcrondal y Rabe-Hesketh (2004).

En los modelos de efectos mixtos el predictor lineal está compuesto por una combinación de efectos fijos y efectos aleatorios. Los efectos fijos son aquellas variables para las cuales el investigador solo ha incluido los niveles que son de su interés. En el caso de un efecto fijo, el interés es comparar los resultados de la variable dependiente para los distintos niveles de la variable explicativa.

Por otra parte, una cantidad es considerada aleatoria cuando cambia sobre las unidades de la población. Cuando se incluye una variable como efecto aleatorio en el modelo, se asume que se busca extraer conclusiones sobre la población sobre la cual se han elegido las unidades observadas y no se tiene interés en esas unidades en particular. Se podría

intercambiar una unidad de la muestra por otra de la población y sería indiferente. En el caso de los efectos aleatorios, el interés no está en la diferencia de medias, sino en cómo el efecto aleatorio explica la variabilidad de la variable dependiente.

Los efectos aleatorios pueden ser incorporados como ordenadas aleatorias o como coeficientes (*pendientes*) aleatorios. Las ordenadas aleatorias representan la heterogeneidad no observada sobre la variable respuesta como conjunto, en tanto los coeficientes aleatorios representan la heterogeneidad no observada en los efectos de las variables explicativas sobre la variable respuesta.

En síntesis, los modelos lineales generalizados mixtos requieren: definir correctamente el predictor lineal, incluyendo las interacciones relevantes; una función de enlace adecuada; correcta especificación de covariables que tendrán coeficientes aleatorios; independencia condicional de las respuestas dados los efectos aleatorios y las covariables; independencia entre los efectos aleatorios y las covariables; y efectos aleatorios con distribución normal. (Rabe-Hesketh y Skrondal, 2012)

Debido que en este trabajo la variable respuesta analizada es binaria (1. Innova, 0. No innova) y los datos presentan una estructura de agrupamiento ya que se trata de medidas repetidas en el tiempo para cada empresa, el modelo particular que se aplica es un **modelo logístico mixto con ordenada al origen aleatoria**.

Sea y_{ij} la variable que representa el estado (1. Innova, 0. No innova) de la empresa i en el momento j , donde $i = 1, \dots, q$ y $j = 1, \dots, t_i$. Condicionado sobre los efectos aleatorios α_i , se supone que y_{ij} son independientes:

$$y_{ij} | \alpha_i \sim \text{Bernulli}(\pi_{ij})$$

$$\text{logit} \{P(y_{ij} = 1 | \mathbf{x}_{ij}, \alpha_i)\} = \ln \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \alpha_i \quad (3.8)$$

donde $P(y_{ij} = 1 | \mathbf{x}_{ij}, \alpha_i) = \pi_{ij}$; el vector \mathbf{x}_{ij} está compuesto por las covariables conocidas; $\boldsymbol{\beta}$ es el vector de parámetros de regresión que representa los efectos fijos de orden $(p \times 1)$,

siendo p la cantidad de covariables; α_i es la ordenada al origen aleatoria con distribución normal con media 0 y varianza ψ^2 . Las ordenadas al origen aleatorias se suponen independientes e idénticamente distribuidas a través de las empresas e independientes de las covariables x_{ij} .

Por su parte, ψ^2 representa el grado de heterogeneidad en las respuestas a través de los sujetos, no atribuible a las covariables. La ordenada aleatoria puede ser pensada como la combinación de los efectos del conjunto de covariables sujeto-específicas omitidas que generan que algunas empresas sean más propensas a innovar que otras. (Rabe-Hesketh y Skrondal, 2012)

Los coeficientes de regresión en (3.8) representan efectos condicionales de las covariables, dados los valores de los efectos aleatorios (sujeto-específico). La esperanza condicional está dada por:

$$E(y_{ij} = 1 | \alpha_i) = \pi_{ij} = \frac{\exp\{x'_{ij}\beta + \alpha_i\}}{1 + \exp\{x'_{ij}\beta + \alpha_i\}} \quad (3.9)$$

Los *odd ratio* obtenidos de los coeficientes de regresión deben ser interpretados condicionados sobre la ordenada aleatoria por lo cual usualmente son denominados *odds ratio* condicionales o sujeto-específicos.

Para realizar inferencia sobre los parámetros del modelo pueden utilizarse las pruebas de razón de verosimilitud (*LR*) y el *test de Wald*. La prueba de razón de verosimilitud utiliza las funciones de máxima verosimilitud estimadas para dos modelos: uno bajo hipótesis nula (l_0) y otro bajo hipótesis alternativa (l_1), de forma tal que:

$$LR = -2 \log\left(\frac{l_0}{l_1}\right) = -2(L_0 - L_1) \sim \chi^2_1$$

Por su parte el estadístico de Wald, para un parámetro individual puede expresarse como:

$$Wald = \frac{\hat{\beta}_i - \hat{\beta}_{i,0}}{\sqrt{\widehat{var}(\hat{\beta}_i)}} \sim AN(0,1)$$

Para realizar selección de modelos, la prueba LR nos permite compara modelos anidados. Pero, ampliando el conjunto de parámetros podemos aumentar la maximización de verosimilitud, aunque un modelo con más parámetros no es necesariamente mejor. El Criterio de información de Akaike (AIC) corrige este sesgo ajustando por el número de parámetros libres, y nos permite compara modelos no anidados. Aquel modelo que presente menor AIC es considerado el mejor.

$$AIC = -2 \log(L) + 2p$$

Donde, L es máxima verosimilitud y p el número de parámetros estimados en el modelo.

3.3.2 Predicción de la variable respuesta en un modelo logístico mixto

En estos modelos los efectos aleatorios son estimados individualmente para cada grupo/sujeto por lo que no es posible hacer predicción directamente para el caso de nuevas unidades, debido a que no se conocen esos valores. A continuación, se presentan una serie de métodos que fueron adaptados por Tamura y Giampaoli (2011 y 2013) y Tamura, *et al.* (2013) para el modelo logístico mixto, con el objetivo de predecir el efecto aleatorio de nuevas unidades/grupos que no formaron parte de la muestra con la que se ajustó el modelo. Para un nuevo grupo l , α_{ml} es desconocido y lo simbolizamos como α_{ml}^* , donde $l=1, \dots, q'$ indica el nuevo grupo/unidad que no fue parte del ajuste del modelo, $m=1, \dots, k$ indica el efecto aleatorio.

Estos métodos son aplicados en el presente trabajo sobre el conjunto de empresas que conforman la muestra de validación (*ver apartado 3.2*), para luego poder analizar y comparar su desempeño. Los métodos de estimación fueron desarrollados para k efectos aleatorios, siendo en esta investigación aplicados a un modelo de un efecto aleatorio (ordenada al origen), tal que $k=1$.

3.3.2.1 Naive

El método más simple es el que supone que el valor del efecto aleatorio para el nuevo grupo es la media de la distribución de los efectos aleatorios conocidos. Este método (*naive*) ignora la existencia de la parte aleatoria, basado en el supuesto del modelo de que $\alpha_1, \dots, \alpha_q$ son *i.i.d* y $\alpha_i \sim N_k(0, \Sigma)$. Por lo tanto, la predicción para nuevos individuos/grupos se realiza considerando solo los efectos fijos estimados en el modelo logístico mixto. La probabilidad predicha para la observación j de un nuevo grupo l , estará dada por:

$$\pi_{lj}^* = P(y_{lj} = 1 | \alpha_l = 0) = \frac{\exp\{\mathbf{x}_{lj}^t \hat{\boldsymbol{\beta}}\}}{1 + \exp\{\mathbf{x}_{lj}^t \hat{\boldsymbol{\beta}}\}} \quad (3.10)$$

La diferencia con el modelo logístico tradicional es que para realizar la predicción para nuevos grupos los parámetros estimados utilizados fueron obtenidos a partir de un modelo con efectos aleatorios.

3.3.2.2 Método del “mejor predictor empírico” (EBP)

Para predecir la respuesta a nivel de observación perteneciente a un nuevo grupo, en el marco de un modelo logístico mixto, basados en el trabajo de Jian y Lahiri (2006), Tamura y Giampaoli (2010) utilizan el mejor predictor empírico. En ambos artículos los autores consideran un modelo solo con constante aleatoria. Tamura y Giampaoli (2011) extienden la propuesta para cuando en el modelo hay k efectos aleatorios. Para resolver las integrales multidimensionales que se plantean, son necesarios métodos de integración numérica.

El Predictor Empírico Bayesiano, dentro del problema considerado, es la media de la distribución empírica posterior, es decir con los parámetros sustituidos por sus estimaciones, $\hat{\zeta} = E(\zeta | y)$. Tiene la propiedad de minimizar el error cuadrático medio de predicción $E(\zeta' - \zeta)^2$ para el predictor ζ' de ζ sobre la distribución conjunta de (ζ) y las respuestas.

Con el objetivo de predecir el valor de la variable respuesta para una nueva observación, definimos a ζ como el resultado del modelo de regresión logística, función que estamos interesados en predecir. De este modo:

$$\zeta(\beta, \alpha_i) = \pi_{ij} = \frac{\exp(x_{ij}^t \beta + z_{ij}^t \alpha_i)}{1 + \exp(x_{ij}^t \beta + z_{ij}^t \alpha_i)} \quad (3.11)$$

Considerando el supuesto del modelo de que $\alpha_i \sim N_k(0, \Sigma)$, para predecir el valor de α_i , de un nuevo grupo l , se utiliza la transformación lineal multivariada $\alpha_i = \Sigma^{1/2} \xi$ donde $\xi \sim N_k(0, I)$. Los autores, haciendo uso de esta transformación, proponen la expresión del mejor predictor empírico (EBP) para predecir la probabilidad de una observación j perteneciente a un nuevo grupo l donde $l \notin Q$:

$$\hat{\pi}_{lj}(\hat{\beta}, \hat{\Sigma}, \xi) = \frac{\exp(x_{ij}^t \hat{\beta}) \int_{\xi_1} \dots \int_{\xi_k} \frac{\exp(y_{l+1} z_{ij}^t \alpha^*)}{1 + \exp(x_{ij}^t \hat{\beta} + z_{ij}^t \alpha^*)} \prod_{l=1}^t \frac{1}{1 + \exp(x_{ij}^t \hat{\beta} + z_{ij}^t \alpha^*)} f(\xi_1, \dots, \xi_k) d\xi_1, \dots, \xi_k}{\int_{\xi_1} \dots \int_{\xi_k} \exp(y_l z_{ij}^t \alpha^*) \prod_{l=1}^t \frac{1}{1 + \exp(x_{ij}^t \hat{\beta} + z_{ij}^t \alpha^*)} f(\xi_1, \dots, \xi_k) d\xi_1, \dots, \xi_k} \quad (3.12)$$

Donde $\hat{\beta}$ es estimado en el modelo (3.8), $f(\xi_1, \dots, \xi_k) = f(\xi_1) \dots f(\xi_k)$ y $f(\xi_m)$ es la función de densidad de una normal estándar univariada con $m=1, \dots, k$. Para un nuevo grupo l , α_i es desconocido y lo simbolizamos como α^* , donde $\alpha^* = \hat{\Sigma}^{1/2} \xi$, con $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_k^*)^t$ tal que:

$$\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_k^* \end{pmatrix} = \begin{pmatrix} \xi_1 \hat{v}_{11} + \xi_2 \hat{v}_{12} + \dots + \xi_k \hat{v}_{1k} \\ \xi_1 \hat{v}_{21} + \xi_2 \hat{v}_{22} + \dots + \xi_k \hat{v}_{2k} \\ \vdots \\ \xi_1 \hat{v}_{k1} + \xi_2 \hat{v}_{k2} + \dots + \xi_k \hat{v}_{kk} \end{pmatrix} \quad (3.13)$$

Con $\hat{v}_{mm'}$, tal que $m = 1, \dots, k$ y $m' = 1, \dots, k$, son los componentes de varianza de la matriz $\hat{\Sigma}^{1/2}$ estimada en el modelo (3.8).

Es importante notar que se busca predecir la variable respuesta, por lo que no se conoce el verdadero valor de y_{li} , es decir cuántas observaciones del grupo l presentaron el valor 1. Se supone a priori que $y_{li} = t_i/2$, lo que implica una probabilidad igual de éxito que de fracaso en el grupo l . Para analizar la influencia de este supuesto sobre los resultados, tomando diferentes valores de y_{li} , mediante estudios de simulación, Giampaoli, *et al.* (2016),

concluyen que la posición de las probabilidades predichas de las observaciones no depende de y_i .

Para resolver las integrales en (3.12) es necesario evaluarlas numéricamente con métodos apropiados de integración multidimensional. En el caso de que haya más de un efecto aleatorio en el modelo, y gran cantidad de observaciones en nuevos grupos sobre los cuales hacer predicción, la aplicación de este método tiene un alto costo computacional. Se requiere de mucho tiempo de procesamiento para resolver las integrales multidimensionales.

3.3.2.3 Método de regresión lineal

El método consiste en ajustar un modelo de regresión lineal que refleje la relación entre un conjunto de variables independientes y los efectos aleatorios estimados para los grupos de la base de entrenamiento, considerando estos últimos como variable respuesta. Luego los parámetros estimados de esta regresión pueden ser usados para predecir los efectos aleatorios de los nuevos grupos.

No se tiene interés en inferir o interpretar los parámetros del modelo lineal, sino que se busca una regla para predecir los efectos aleatorios desconocidos, previo cumplimiento de los supuestos y evaluación sobre las variables que resultan significativas.

Siguiendo a Tamura y Giampaoli (2013) primero es necesario agregar a nivel de grupo/empresa (i) las variables disponibles a nivel de observación (j), ya que la regresión es ajustada a nivel de grupo. Hay numerosas maneras de agregar las variables. Se pueden usar funciones como la media, la mediana, el mínimo, el máximo, entre otros. La elección de la función a utilizar para sintetizar la información dependerá del problema analizado. La misma variable puede ser testeada utilizando diferentes criterios. Luego se ajusta una regresión lineal para cada efecto aleatorio estimado en el modelo mixto (3.8), considerando los grupos de la base de entrenamiento.

$$\hat{\alpha}_{mi} = \mathbf{W}_{mi}^t \lambda_m + \varepsilon_{mi} \quad \varepsilon_{mi} \sim N(0, \sigma_m^2) \quad (3.14)$$

Donde:

$i = 1, \dots, q$ indica el grupo, tal que $i \in Q = \{1, \dots, q\}$; unidades de la base de entrenamiento,

$m = 1, \dots, k$ indica el efecto aleatorio,

\mathbf{W}_{mi} es el vector de variables agregadas a nivel de grupo

Con las covariables seleccionadas y los parámetros estimados en el paso anterior, se predicen los efectos aleatorios del nuevo grupo l de la base de validación.

$$\hat{\alpha}_{ml}^* = \mathbf{W}_{ml}^t \hat{\lambda}_m \quad \text{tal que } l \notin Q \quad (3.15)$$

Finalmente, utilizando el modelo logístico mixto, es posible predecir la probabilidad de que una observación perteneciente a un nuevo grupo presente el fenómeno estudiado, considerando los valores de los efectos aleatorios predichos en (3.15) y los efectos fijos estimados en el modelo.

$$\hat{\pi}_{lj}^* = \frac{\exp\{x_{lj}^t \hat{\beta} + z_{lj}^t \hat{\alpha}_l^*\}}{1 + \exp\{x_{lj}^t \hat{\beta} + z_{lj}^t \hat{\alpha}_l^*\}} \quad (3.16)$$

3.3.2.4 Método del vecino más cercano

Este método es una técnica no paramétrica, comúnmente utilizada en clasificación supervisada. Siguiendo a Tamura, *et al.* (2013), para realizar la predicción de los efectos aleatorios de los nuevos grupos se consideran los valores conocidos de estos efectos de sus vecinos más cercanos. Al ser el valor de los efectos aleatorios una variable continua es necesario considerar alguna medida de centralidad (media, mediana, u otra) para realizar la imputación del valor correspondiente. Se selecciona la cantidad de vecinos a considerar de manera que se maximice la performance de predicción del modelo mixto. La ventaja del método es que no requiere de ninguna distribución para los efectos aleatorios.

En primera instancia se calculan las distancias (Euclídea, Mahalanobis, Minkowski, City Block, u otra) entre los nuevos grupos (base de validación) y aquellos con efectos aleatorios conocidos (base de entrenamiento), considerando su vector de características \mathbf{g} .

Para $i'=1\dots q'$ donde $i' \notin Q$ (grupos de la base de validación),

$i=1\dots q$ donde $i \in Q$ (grupos de la base de entrenamiento),

se computan las distancias $d_{(i', i)}$ entre $\mathbf{g}_{i'}$ y \mathbf{g}_i , y luego se ordenan las distancias de manera creciente,

$$d_{(i', .)} = (d_{(i', 1)}, d_{(i', 2)}, \dots, d_{(i', q)})$$

Posteriormente se computan medidas de centralidad para los efectos aleatorios conocidos correspondientes a los l primeros elementos de las distancias ordenadas.

$$\hat{\alpha}_{mi'}^* = (\hat{\alpha}_{m1}, \hat{\alpha}_{m2}, \dots, \hat{\alpha}_{ml})$$

Finalmente, se introducen los efectos aleatorios en el predictor lineal de la regresión logística mixta, como en 3.16, haciendo posible la predicción de la probabilidad de que una observación de un nuevo grupo presente la característica estudiada.

3.3.2.5 Comparación de los métodos de predicción: estudios empíricos

Los métodos de predicción de los efectos aleatorios presentados han sido desarrollados y aplicados a diferentes problemas de la realidad (Tabla 3.3). En la investigación de Tamura y Giampaoli (2013) se trabaja sobre un problema de marketing en compañías de telecomunicaciones, donde se quiere predecir la probabilidad de que nuevos clientes adquieran un servicio de la empresa. Comparan el método naive, con el de EBP y el de regresión lineal concluyendo que éstos últimos mejoran las medidas de predicción del modelo. Los autores destacan que ambos métodos resultan robustos a la falta de normalidad en la distribución de los efectos aleatorios, y advierten sobre el menor costo

computacional del método de regresión lineal con relación al método de EBP, situación relevante en su caso donde trabajan con cerca de 20.000 observaciones.

En el trabajo de Tamura, *et al.* (2013), en el cual se propone la estimación de los efectos aleatorios mediante el método del vecino más cercano, se compara éste con el EBP y naive en un ejercicio de aplicación sobre información nutricional en niños. Se concluye que, con pequeñas diferencias, se logran mejores resultados con el nuevo método propuesto.

Los trabajos de Giampaoli, *et al.* (2016) y Caro, *et al.* (2017), comparan los tres métodos presentados con el naive, en un problema de predicción de crisis financiera de empresas latinoamericanas. En el primer trabajo se arriba a la conclusión de que el método del vecino más cercano es el que arroja mejores resultados, seguido por naive, el de regresión lineal y finalmente el de EBP. En la segunda investigación se estudia la predicción de crisis financiera de las empresas de cada país (Argentina, Chile y Perú) por separado. El método de regresión lineal no resultó apropiado para el problema ya que las covariables disponibles no resultaron significativas para explicar el efecto aleatorio de las nuevas empresas. El del vecino más cercano mejoró la precisión del modelo respecto al naive sólo en el caso de argentina.

Tabla 3.3: Estudios empíricos seleccionados sobre predicción de efectos aleatorios

Autor/es	Año	Problema de aplicación	Métodos comparados	Principales conclusiones
Tamura K. y Giampaoli V.	2013	Predicción de la probabilidad de que nuevos clientes adquieran un servicio de la empresa (marketing - telecomunicaciones)	Naive, EBP, Regresión lineal	Ambos metodos producen mejoras frente al Naive, con diferencias pequeñas entre ellos, siendo robustos a la falta de normalidad de los efectos aleatorios. El método de Regresión lineal tiene menor costo computacional.
Tamura K., Giampaoli V. y Noma, A.	2013	Clasificación en base a información nutricional en niños	EBP, Vecino más cercano	El método del Vecino más cercano produce mejores resultados que el de EBP
Giampaoli V., Tamura K., Caro N., Simoes de Araujo L.	2016	Predicción de crisis financiera de empresas latinoamericanas (Argentina, Chile y Peru)	Naive, EBP, Regresión lineal, Vecino más cercano	Pequeñas diferencias entre los métodos. El mejor resultado fue el del Vecino más cercano, seguido por Naive, Regresión lineal y finalmente por EBP.
Caro N., Arias V., Ortiz P.	2017	Predicción de crisis financiera de empresas latinoamericanas (Argentina, Chile y Peru). Análisis de los países por separado	Naive, Regresión lineal, Vecino más cercano	Método de Regresión lineal no resultó adecuado. Solo para Argentina el método del Vecino más cercano mejora los resultados respecto al Naive.

Fuente: Elaboración propia.

De la revisión de los trabajos de aplicación de los distintos métodos de predicción de los efectos aleatorios, surge que las diferencias de sus efectos sobre las medidas de desempeño del modelo son pequeñas. En aquellos trabajos en que se evaluó el método del vecino más cercano éste fue el que mejores resultados generó (Tamura, *et al.*, 2013; Giampaoli, *et al.*, 2016; Caro, *et al.*, 2017). Respecto a los métodos EBP y de regresión lineal no se encontraron resultados concluyentes.

3.3.3 Medidas de desempeño del modelo

Cuando se construye un modelo con el objetivo de hacer predicción es muy importante su desempeño, tanto sobre la base de datos utilizada para la construcción del modelo (base de entrenamiento) como sobre una base de predicción (base de validación). Para evaluar el poder predictivo del modelo se utilizan herramientas que permiten comparar los valores observados de la variable respuesta con sus estimaciones.

Cada valor de probabilidad estimada de ocurrencia del evento ($\hat{\pi}_{ij}$), puede ser transformado en binario (éxito/fracaso) según un punto de corte o umbral (c) establecido, cuya elección dependerá de cada problema analizado.

De este modo si: $\hat{\pi}_{ij} \geq c ; \hat{y}_{ij} = 1$

$$\hat{\pi}_{ij} < c ; \hat{y}_{ij} = 0$$

La tabla 3.4, llamada tabla de clasificación muestra los valores observados de la variable respuesta versus la clasificación predicha por el modelo en el punto de corte definido.

Tabla 3.4: Tabla de clasificación

		Valor Predicho (modelo)		Total
		1	0	
Valor Observado	1	VP	FN	VP + FN
	0	FP	VN	FP + VN
Total		VP + FP	FN + VN	VP+FN+FP+VN

De esta forma, surgen observaciones que presentando efectivamente el evento de interés son clasificadas adecuadamente por el modelo, verdaderos positivos (VP), y aquellas que son clasificadas incorrectamente, es decir falsos negativos (FN). Por otra parte, las observaciones que no poseen la característica estudiada pueden ser clasificadas correctamente, verdaderos negativos (VN) o incorrectamente, falsos positivos (FP). A través de la comparación de las predicciones del modelo con las respuestas observadas en la muestra se pueden calcular medidas de acierto y tasas de error.

Una medida de precisión general del modelo es la proporción de observaciones clasificadas correctamente (*pcc*), es decir:

$$pcc = \frac{VP + VN}{Total\ de\ observaciones}$$

La sensibilidad mide la proporción de aciertos en el evento de interés, e indica que tan bueno es el modelo para identificar las observaciones con respuesta igual a 1.

$$sensibilidad = \frac{VP}{FN + VP}$$

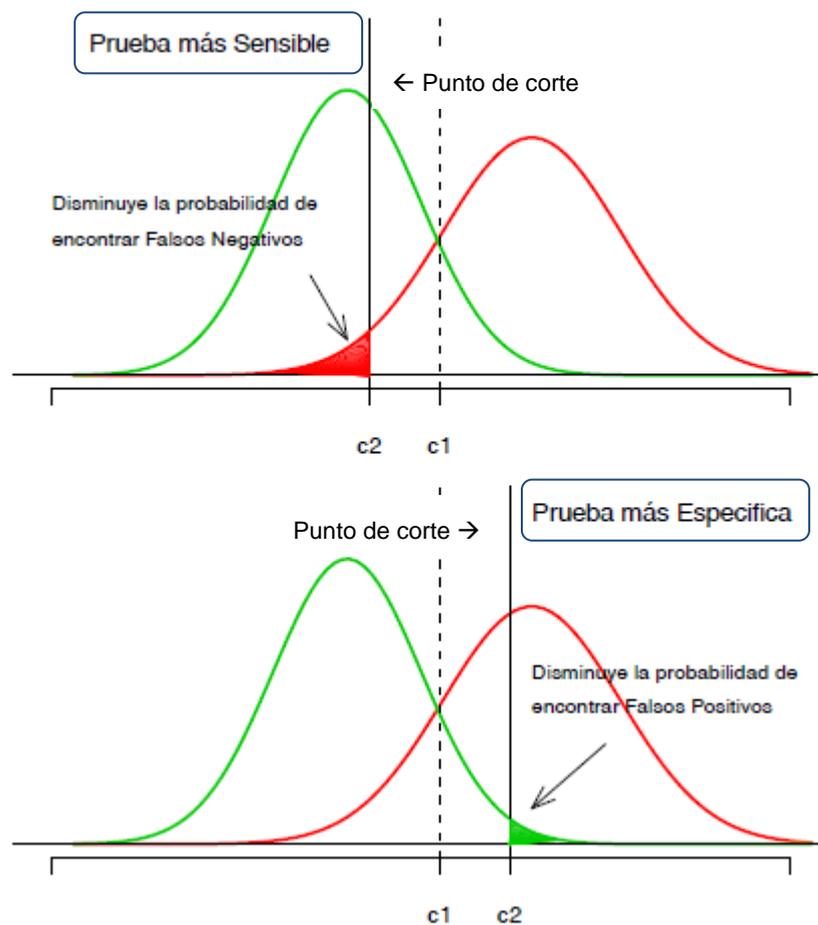
En tanto, la especificidad indica que tan bueno es el modelo para identificar las observaciones con respuesta igual a 0.

$$especificidad = \frac{VN}{VN + FP}$$

Estas medidas varían entre 0 y 1, por lo que un modelo con buenas predicciones presenta valores cercanos a 1.

Las medidas descriptas dependen del umbral o punto de corte elegido (c). Las unidades que no presentan la característica estudiada y aquellas que si la presentan siguen distribuciones diferentes, originando hipotéticamente las curva (verde y rojo respectivamente) que se muestra en el gráfico 3.1. Resultan deseables modelos con alta sensibilidad y especificidad, pero al reducir el punto de corte el aumento de sensibilidad se hace en detrimento de la especificidad y viceversa.

Gráfico 3.1: Efecto de modificaciones en el punto de corte



La definición del punto de corte puede responder a diversos criterios en función del problema analizado. Una alternativa es determinar el umbral como aquel que iguala el nivel de sensibilidad con el de especificidad, o aquel que maximiza la proporción de clasificación correcta. Otra opción es definir el punto de corte tal que la distancia entre la curva ROC (*Receiver Operating Characteristic*) y el punto (0,1) sea mínima.

La curva ROC es un gráfico que permite observar todos los pares de sensibilidad y complemento de la especificidad, resultantes de la variación del punto de corte en todo el rango de resultados observados (entre 0 y 1 en nuestro caso).

$$ROC(.) = \{sensibilidad(c), 1 - especificidad(c), c \in (0,1)\}$$

Para puntos de corte cercanos a 1 ambas medidas se aproximan a cero, por el contrario, cuando c está cerca de 0 las medidas tienden a 1. Así, la curva ROC es una función monótona creciente en el cuadrante positivo.

Para un nivel dado de especificidad, se prefiere el modelo que tenga mayor sensibilidad, es decir cuya curva ROC sea más alta. El área bajo la curva (AUC), estima la capacidad del modelo de “discriminar” entre quienes presentan la característica bajo estudio y quiénes no. Si el área bajo la curva tomara el valor 1 (la curva se ubica sobre la esquina superior izquierda), la predicción del modelo sería perfecta. Si en cambio, el área bajo la curva tomara el valor 0,5, el modelo sería no informativo. Existiría la misma probabilidad de clasificar a quien presenta la característica estudiada como tal, o como si no la presentara.

El uso de las curvas ROC en la evaluación de modelos de clasificación o pruebas diagnósticas presenta las ventajas de ser curvas fáciles de interpretar, no requieren un nivel de decisión particular porque está incluido todo el espectro de puntos de corte y proporciona una comparación visual directa entre modelos (Zweig y Campbell, 1993).

Las medidas aquí presentadas serán utilizadas en la aplicación sobre el problema de innovación en empresas industriales de la provincia de Córdoba para la validación del modelo ajustado y para la comparación de las diferentes alternativas utilizadas en la predicción del efecto aleatorio sobre nuevas unidades,

Capítulo 4: Resultados

En este capítulo, se presentan los resultados encontrados en esta investigación. En primer lugar, se realiza un análisis exploratorio de la muestra de empresas utilizadas en la presente investigación, observando su evolución y particularidades del proceso de innovación. Luego, se presentan los resultados de la aplicación del modelo logístico mixto, en el que se evalúa la influencia de un conjunto de factores sobre la probabilidad de obtener mejoras significativas y/o nuevos productos y/o procesos, novedosos para el mercado. Finalmente, se compara la performance de los métodos de predicción para los efectos aleatorios presentados en el capítulo 3, a fin de predecir la probabilidad de innovar de empresas que no formaron parte de la muestra con la que se estimó el modelo. Todo el análisis fue realizado utilizando el software R versión 3.4.4².

4.1 Innovación en la provincia de Córdoba

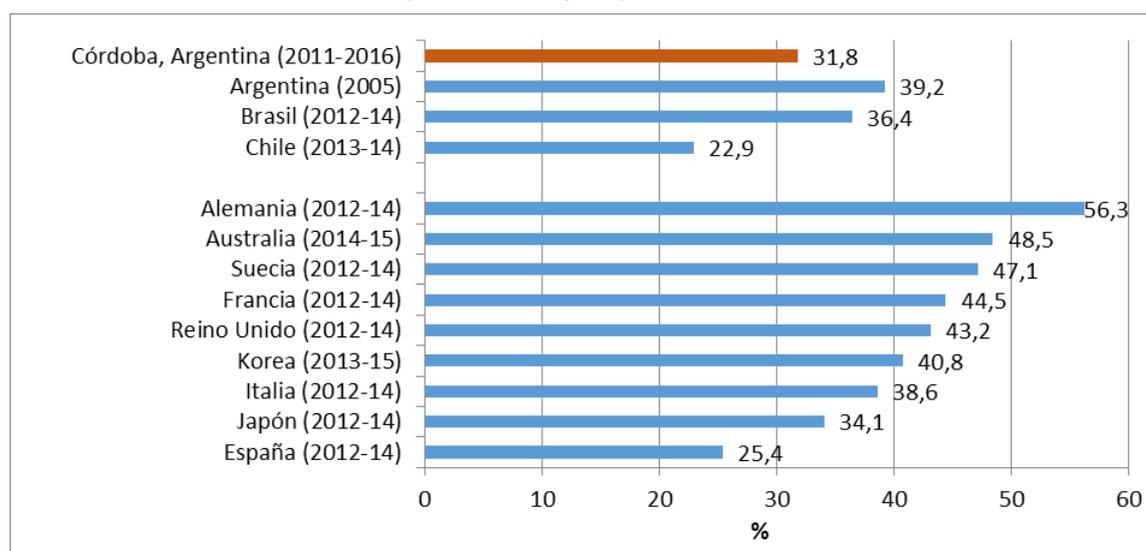
La provincia de Córdoba, localizada en la región centro de Argentina, tiene una población de 3.308.876 habitantes según el censo nacional 2010, con lo cual es la segunda provincia más poblada del país. Sus principales actividades económicas en el período bajo análisis son la actividad agropecuaria y la actividad industrial. La participación promedio de la industria manufacturera en el producto bruto provincial, a valores constantes de 2004, fue de 19%, por lo que es de gran relevancia el estudio de la innovación en las empresas del sector.

Como puede observarse en el gráfico 4.1, en la provincia de Córdoba el índice de innovación en empresas manufactureras, definido como la proporción de empresas que lograron innovaciones en productos y/o procesos (pudiendo haber logrado o no resultados

² R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

en organización y/o comercialización), fue en promedio durante el período analizado de 31,8%. Este valor es superior al 28% estimado para el conjunto de países de América Latina y el Caribe³, y a la proporción de empresas innovadoras en Chile; pero se encuentra por debajo del valor registrado en la mayoría de los países miembros de la OECD.

Gráfico 4.1: Proporción de empresas industriales que lograron innovaciones en productos y/o procesos.

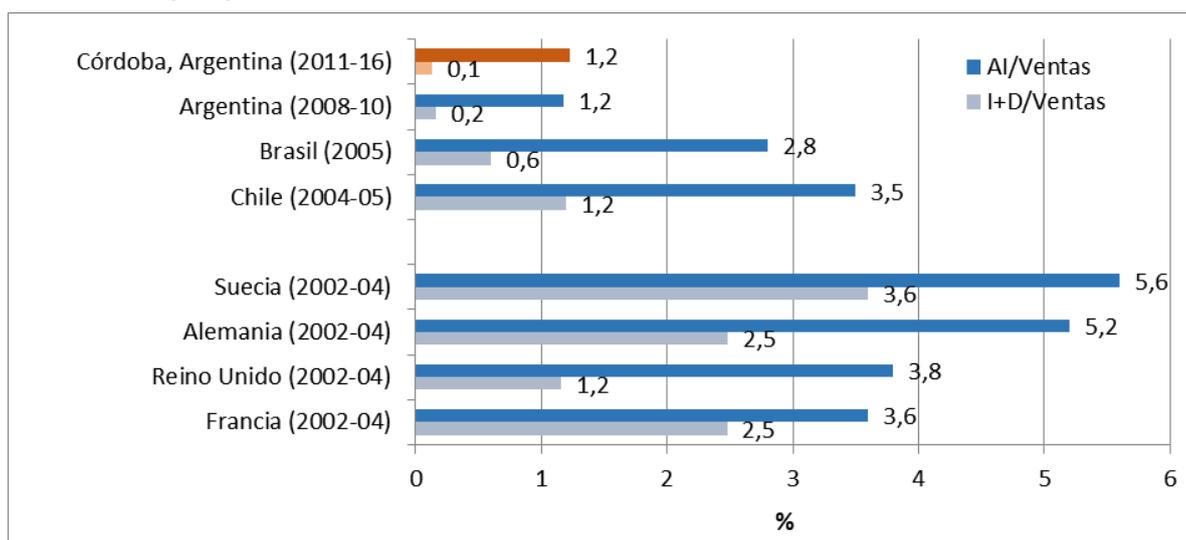


Fuente: Córdoba, Argentina: Encuesta Provincial de Innovación y Conducta Tecnológica 2011-2016, Dir. Gral. de Estadística y Censos Córdoba. Argentina: Encuesta Nacional de Innovación y conducta Tecnológica, Resultados año 2005 – INDEC. Resto de los países: OECD Science, Technology and Industry Scoreboard 2017 - © OECD 2017

La inversión en actividades de innovación realizada por empresas manufactureras de la provincia fue del 1,2% de las ventas, en tanto los gastos en I+D interna fueron el 0,1%, durante el período 2011-2016. A nivel nacional, la inversión en estos conceptos fue de 1,2% y 0,2% de las ventas respectivamente para el período 2008-2010. La comparación internacional coloca a los esfuerzos realizados por las empresas argentinas, y a las localizadas en la provincia de Córdoba en particular, por debajo del promedio de los países europeos, y de sus pares de la región.

³ Arias Ortiz *et al.* (2013), en base a Enterprice Survey 2010 - BID

Gráfico 4.2: Inversión en actividades de innovación (AI) y en investigación y desarrollo (I+D) en relación con las ventas.



Fuente: Córdoba, Argentina: Encuesta Provincial de Innovación y Conducta Tecnológica 2011-2016, Dir. Gral. De Estadística y Censos Córdoba. Argentina: Encuesta Nacional de Innovación y conducta Tecnológica, INDEC. Resto de los países: Crespi y Zúñiga (2010), en base a datos OECD (2009), y encuestas de innovación para Brasil, Chile y Colombia.

4.2 Descripción de la muestra

De las 508 empresas consideradas en la presente investigación, 276 conforman la muestra de entrenamiento utilizada para ajustar el modelo. Las 232 empresas restantes, son parte de la muestra utilizada para la validación de éste y la evaluación de los métodos de predicción de los efectos aleatorios. Las características de las empresas que conforman ambas muestras son similares.

Como puede observarse en la tabla 4.1, en promedio en el período 2011-2016, el 27% de las empresas de la muestra de entrenamiento lograron, como resultado de sus esfuerzos innovadores, productos y/o procesos nuevos y/o con mejoras significativas novedosos para el mercado nacional o internacional. El indicador es similar para todo el período. Alcanza su valor mínimo de 24% en el año 2012, y el máximo de 29% en año 2014. Por su parte, la muestra de validación posee un 29% de empresas innovadoras⁴.

⁴ Estos valores difieren del presentado en el gráfico 4.1 porque se acotó el análisis a pequeñas y medianas empresas (PyMes).

Tabla 4.1: Proporción de empresas innovadoras, promedio del período 2011-2016

		<i>Muestra</i>	
		<i>Entrenamiento</i>	<i>Validación</i>
<i>INNOVA</i>	<i>SI</i>	27%	29%
	<i>NO</i>	73%	71%

Como fue mencionado en la sección 3.2, una de las variables evaluadas en el modelo es el tamaño de las empresas el cual, siguiendo las recomendaciones del manual de Oslo (2005), se mide sobre la base del número de empleados. En la conformación de las muestras de entrenamiento y validación se restringió el análisis a empresas pequeñas y medianas, entendiendo como tal a las que tenían entre 5 y 250 empleados (promedio 2011-2016).

Tabla 4.2: Tamaño de las empresas. Personal ocupado

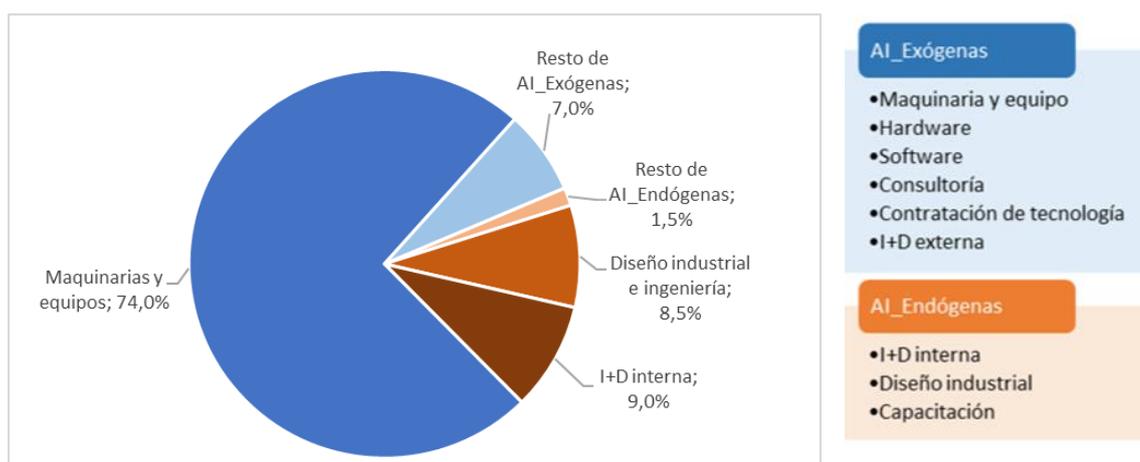
<i>RRHH</i>	<i>Muestra de Entrenamiento</i>		<i>Muestra de Validación</i>	
	<i>INNOVA</i>		<i>INNOVA</i>	
	<i>SI</i>	<i>NO</i>	<i>SI</i>	<i>NO</i>
<i>Media</i>	69,2	45,7	61,3	32,5
<i>Desv. Est.</i>	64,4	55,7	59,8	42,3
<i>CV</i>	0,9	1,2	1,0	1,3
<i>Mediana</i>	38,0	19,0	41,0	16,0
<i>Mínimo</i>	2,0	1,0	4,0	1,0
<i>Máximo</i>	278,0	290,0	251,0	241,0

Como puede observarse en la tabla 4.2, las empresas innovadoras son en promedio de mayor tamaño que las empresas no innovadoras. La mitad de las empresas innovadoras de la muestra de entrenamiento tienen hasta 38 empleados, en tanto las no innovadoras tienen 19 empleados o menos. Las medidas descriptivas de la variable tamaño, según sean innovadoras o no, son similares en ambas muestras (entrenamiento y validación).

Entre las variables predictoras relacionadas a los esfuerzos en materia de innovación realizados por las empresas, se incluyen en el modelo la proporción en relación a las ventas de los gastos en actividades de innovación. Se analizan por separado las actividades endógenas a la empresa, tales como gastos en investigación y desarrollo interna, gastos en capacitación y/o en actividades de diseño industrial e ingeniería; y actividades de innovación exógenas a la empresa, en las cuales se incluyen gastos en investigación y desarrollo externos, adquisición de maquinaria y equipo, de hardware, de software, contratación de tecnología y/o de consultorías.

En el gráfico 4.3 se muestra la composición de los gastos en actividades de innovación (AI). En promedio en el período analizado, el 81% del total invertido fue destinado a actividades de innovación exógenas, y el 19% restante a actividades de innovación endógenas a la empresa. Del gasto total, la inversión en maquinaria y equipos es la actividad principal (74%), seguida por actividades de investigación y desarrollo interna (9%), y por actividades de diseño industrial e ingeniería (8,5%).

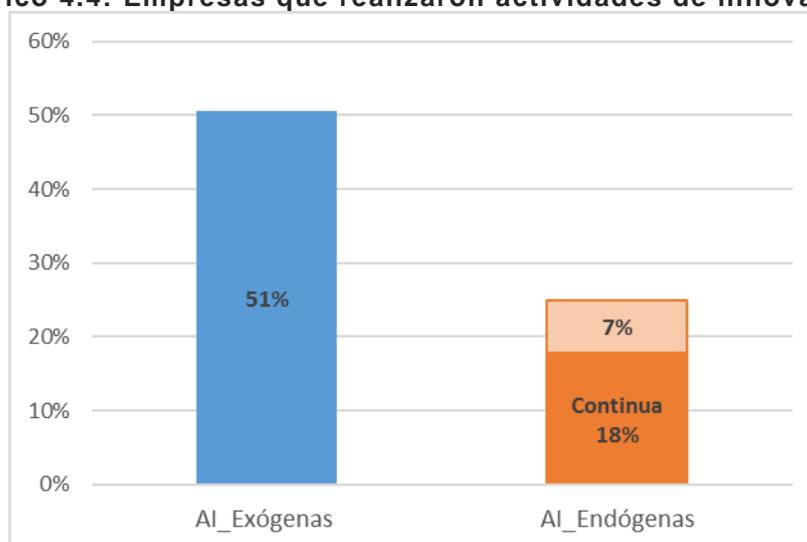
Gráfico 4.3: Composición del gasto en actividades de innovación. Promedio 2011-2016



Una conducta equilibrada en materia de AI implica que la empresa decide invertir en forma pareja en los diversos rubros que componen los esfuerzos innovativos, situación que no se observa en las empresas analizadas. Una estructura de AI más equilibrada, aumenta las probabilidades de que la empresa aproveche más plenamente tanto los esfuerzos endógenos de I+D, ingeniería y diseño y capacitación, como los esfuerzos exógenos (compra de maquinaria, adquisición de tecnología desincorporada, I+D externa, etc) (Lugones *et al*, 2005, RICYT - Manual de Bogotá, 2001).

Del total de empresas que conforman la muestra de entrenamiento, el 51% realizó gastos en AI externas a la empresa y el 25% realizó actividades de innovación intra-muros (gráfico 4.4). Los esfuerzos realizados por las empresas en actividades de innovación pueden no generar resultados inmediatos. El carácter acumulativo del conocimiento hace que sea necesario destinar un flujo constante de recursos hacia estas áreas para obtener resultados exitosos. Es por esto que otra variable considerada en el modelo es la continuidad del gasto en actividades de innovación endógena (4 años o más, en el período bajo análisis). Como puede observarse en el gráfico, del total de empresas consideradas para el ajuste del modelo, el 18% realizó gastos en AI endógenas en forma continua, mostrando estrategias de innovación de largo plazo.

Gráfico 4.4: Empresas que realizaron actividades de innovación.



En la tabla 4.3, se puede observar la proporción de empresas con gastos en AI según si han logrado o no resultados innovadores en productos y/o procesos novedosos para el mercado. El 63% de las empresas innovadoras, de la base de entrenamiento, realizaron gastos en actividades de innovación endógenas, y el 94% realizó gastos en AI exógenas. Por otra parte, el 11% de las empresas que no lograron resultados innovadores en productos y/o procesos realizó gastos en AI endógenas, y el 35% realizó gastos en AI exógenas. Comportamiento similar se observa en las empresas que conforman la base de validación.

Tabla 4.3: Proporción de empresas que realizaron gastos en AI según sean innovadoras o no. Promedio del período 2011-2016

	Muestra de Entrenamiento		Muestra de Validación	
	INNOVA		INNOVA	
	SI	NO	SI	NO
<i>AI_Endógenas (SI)</i>	63%	11%	63%	14%
<i>AI_Exógenas (SI)</i>	94%	35%	94%	29%

Las empresas innovadoras realizaron gastos promedio en actividades de innovación endógena del orden del 1,2% de sus ventas, en tanto las no innovadoras invirtieron en promedio el 0,1% de sus ventas. Por su parte, el 3,7% de las ventas fueron destinados a actividades de innovación exógena en las empresas que lograron resultados innovadores, mientras que las empresas que no lograron innovaciones en productos y/o procesos gastaron en promedio en estas actividades el 1,2% de las ventas (Tabla 4.4).

Tabla 4.4: Promedio de gastos en AI como % de las ventas totales

	Muestra de Entrenamiento		Muestra de Validación	
	INNOVA		INNOVA	
	SI	NO	SI	NO
<i>AI_Endógenas / Vtas</i>	1,2%	0,1%	1,2%	0,1%
<i>AI_Exógenas / Vtas</i>	3,7%	1,2%	4,2%	0,8%

Invertir en actividades de innovación puede implicar un gran esfuerzo financiero por parte de las empresas. Alguno de los obstáculos señalados por las firmas para invertir en I+D interna es el alto riesgo y costo de la misma, como así también la falta de acceso al crédito. Como se muestra en la tabla 4.5, en lo referido al uso de recursos de programas oficiales de promoción de la innovación, el 22% de las empresas innovadoras los han utilizado. De las empresas no innovadoras, el 6% utilizó esa fuente de financiamiento.

Tabla 4.5: Relación con el sistema nacional de innovación

	Muestra de Entrenamiento		Muestra de Validación	
	INNOVA		INNOVA	
	SI	NO	SI	NO
<i>Utilizó fondos de programas oficiales</i>	22%	6%	18%	7%
<i>Actividades de Vinculación</i>	48%	13%	36%	12%

Como ya fue mencionado en el capítulo 2, las firmas no innovan de forma aislada, el proceso de innovación involucra un sistema de interacciones e interdependencias entre las firmas y otras organizaciones e instituciones. Como se observa en la tabla 4.5, el 48% de las empresas innovadoras (muestra de entrenamiento) declaró haber tenido algún tipo de vinculación con entidades públicas o privadas del sistema de innovación nacional (Universidades, INTI, INTA, CONICET, otros organismos gubernamentales, clientes, proveedores, laboratorios, etc); ya sean acuerdos cooperativos con participación activa o intercambio formal o informal de información. Del grupo de empresas que no lograron resultados innovadores, el 13% mantuvo algún tipo de vinculación con estas instituciones.

4.3 Determinantes de la Innovación

4.3.1 Resultados de la estimación

En el presente apartado se exponen los resultados obtenidos en el ajuste del modelo utilizado para identificar y evaluar los principales determinantes de la innovación. Como fue detallado en el capítulo 3, el modelo particular que se aplica en este trabajo es un modelo logístico mixto con ordenada aleatoria (la empresa). La variable respuesta es binaria e indica el grupo al que pertenece la firma (innova= 1 / no innova=0) y las variables predictoras son indicadores vinculados a la estructura de las empresas, a su comportamiento relacionado a los esfuerzos innovadores y a sus vinculaciones con otras entidades del sistema nacional de innovación (detalladas en la tabla 3.2). El modelo a estimar queda expresado de la siguiente manera:

$$\begin{aligned} \text{logit} \{P(y_{ij} = 1 | \mathbf{x}_{ij}, \alpha_i)\} = & \beta_0 + \beta_1 LN_{RRHH} + \beta_2 AI_{ENDOVT_A} + \beta_3 AI_{EXOVT_A} \\ & + \beta_4 AI_{ENDOCONT_A} + \beta_5 FONDOS_{PROGOFIC} \\ & + \beta_6 VINCULA + b_j \end{aligned}$$

Donde:

$P(y_{ij} = 1 | \mathbf{x}_{ij}, \alpha_i) = \pi_{ij}$: probabilidad de éxito en obtener resultados innovadores en productos y/o procesos nuevos o con mejoras significativas, novedosos para el mercado nacional y/o internacional.

b_j : ordenada aleatoria (empresa)

Para el ajuste del modelo se utilizó la función *glmer* disponible en la librería *lme4* del software R. La integral sobre el efecto aleatorio de la función de verosimilitud es aproximada mediante el método de cuadratura adaptativa de Gauss-Hermite.

En la Tabla 4.6. se presentan los coeficientes estimados del modelo. El signo de los coeficientes es el esperado para todas las variables analizadas. Como efecto aleatorio, se consideró la empresa, dado que en algunos períodos ha innovado y en otros no. El valor

de probabilidad asociado a la prueba de razón de verosimilitud (LR) muestra que hay un cambio significativo en la función de log de la verosimilitud al pasar de un modelo restringido (sin efectos aleatorios) al modelo especificado. Es decir, que existe efecto empresa para explicar la mayor proporción de la heterogeneidad inducida por los datos, lo que justifica su inclusión como coeficiente aleatorio.

Tabla 4.6: Estimación de los parámetros del modelo logístico mixto

<i>Efectos Fijos</i>	<i>Coefficiente</i>	<i>Error Estándar</i>	<i>p-value</i>	<i>Exp (β)</i>
<i>LN_RRHH</i>	0,478	0,118	< 0,001	1,613
<i>AI_ENDO_VTA</i>	0,714	0,125	< 0,001	2,042
<i>AI_EXO_VTA</i>	0,060	0,018	< 0,001	1,062
<i>AI_ENDO_CONT4</i>	2,011	0,394	< 0,001	7,469
<i>FONDOS_PROGOFIC</i>	0,672	0,310	0,030	1,958
<i>VINCULA</i>	1,118	0,230	< 0,001	3,060
<i>Constante</i>	-4,393	0,458	< 0,001	
<i>Efecto Aleatorio</i>	<i>Varianza</i>	<i>Error Estándar</i>	<i>LR test vs. Regresión logística Pr(Chibar2)</i>	
<i>EMPRESA</i>	2,884	1,698	0,000	

El tamaño de la empresa, medido con la cantidad de personal ocupado (en logaritmo), resultó un factor significativo relacionado positivamente con la probabilidad de la firma de innovar en productos y/o procesos. Ante un incremento unitario en el tamaño de la empresa medido en logaritmo, la chance de innovar aumenta en un 61,3%. Este resultado es respaldado con lo observado en otros trabajos realizados para nuestro país (Chudnovsky *et al.*, 2004; Crespi y Zúñiga 2010; Milesi, *et al.*, 2011), y se encuentra en línea con la hipótesis schumpeteriana según la cual son mayores las posibilidades innovadoras en el caso de grandes firmas y con poder de mercado.

En un contexto de debilidad en el funcionamiento de los mercados e instituciones, las firmas grandes están mejor posicionadas para sobreponerse a las restricciones del contexto.

Crespi y Zúñiga (2010), destacan que las empresas de mayor tamaño tienen un mejor posicionamiento para aprovechar inversiones en innovación, producto de las economías de escala relacionadas a la producción y a la diversificación de I+D, el mayor acceso al financiamiento y una mejor apropiación de la difusión de conocimientos externos.

Las variables incluidas en el modelo relacionadas al esfuerzo en actividades de innovación realizado por las empresas, tanto endógenas como exógenas, resultaron estadísticamente significativas. Aquellas empresas que invierten una mayor proporción de sus ventas en actividades de innovación poseen mayor probabilidad de obtener productos y/o procesos novedosos para el mercado, siendo de mayor impacto las actividades de innovación internas a la empresa. Evidencias de esta relación positiva entre gastos en actividades de innovación y resultados innovadores pueden encontrarse tanto en trabajos realizados para países desarrollados (Griffith *et al.*, 2006; Mairesse y Mohnen, 2010) como en desarrollo (Chudnovsky *et al.*, 2004; Benavente, 2005; Raffo *et al.*, 2008; Arza y López, 2010; Marín *et al.*, 2017 y Gómez y Borrastero, 2018).

Un incremento unitario en la inversión en actividades de innovación endógenas (I+D interna, diseño e ingeniería, capacitación) con relación a las ventas, genera un aumento en la chance de innovar del 104,2%. En tanto si el incremento unitario se produce en las actividades exógenas (adquisición de maquinaria y equipos, hardware, software, consultoría, contratación de tecnología, I+D externa) en relación a las ventas, la chance de innovar aumenta 6,2%.

Chudnovsky *et al.* (2004), y Marín, *et al.* (2017) destacan en sus trabajos la importancia de que las firmas aprendan al innovar y que dicho proceso de aprendizaje debe ser continuo para ser efectivo. Respecto a la continuidad del gasto en AI endógenas, el coeficiente estimado en el modelo es significativo y con signo positivo. La chance de lograr innovaciones en productos y/o procesos novedosos para el mercado, es más de seis veces superior en una firma que gasta en actividades de innovación internas a la empresa de

forma continua comparada con aquellas que lo hacen de manera esporádica. De esta forma, la continuidad en el gasto en AI endógenas se destaca como la variable de mayor influencia sobre la probabilidad de obtener resultados innovadores. Las empresas que gastan en I+D en forma sistemática, son las que más aprovechan los efectos acumulativos y de aprendizaje derivados de la continuidad de la investigación tecnológica (Buesa *et al.*, 2002).

Las características propias de la inversión en innovación implican grandes esfuerzos por parte de las empresas para llevarlas a cabo. Tal como plantea Crespi y Zúñiga (2010), una de las principales dificultades percibidas por las firmas es la de esperar largos períodos de tiempo para ver resultados de la inversión, como así también los altos costos y riesgos que tienen asociadas las actividades de innovación. En el modelo, la relación entre el uso de fondos de programas públicos y la obtención de resultados innovadores es estadísticamente significativa. Aquellas firmas que han utilizado fondos de programas oficiales de fomento a la innovación tienen una chance un 95,8% superior de obtener resultados innovadores, que aquellas que no los han utilizado.

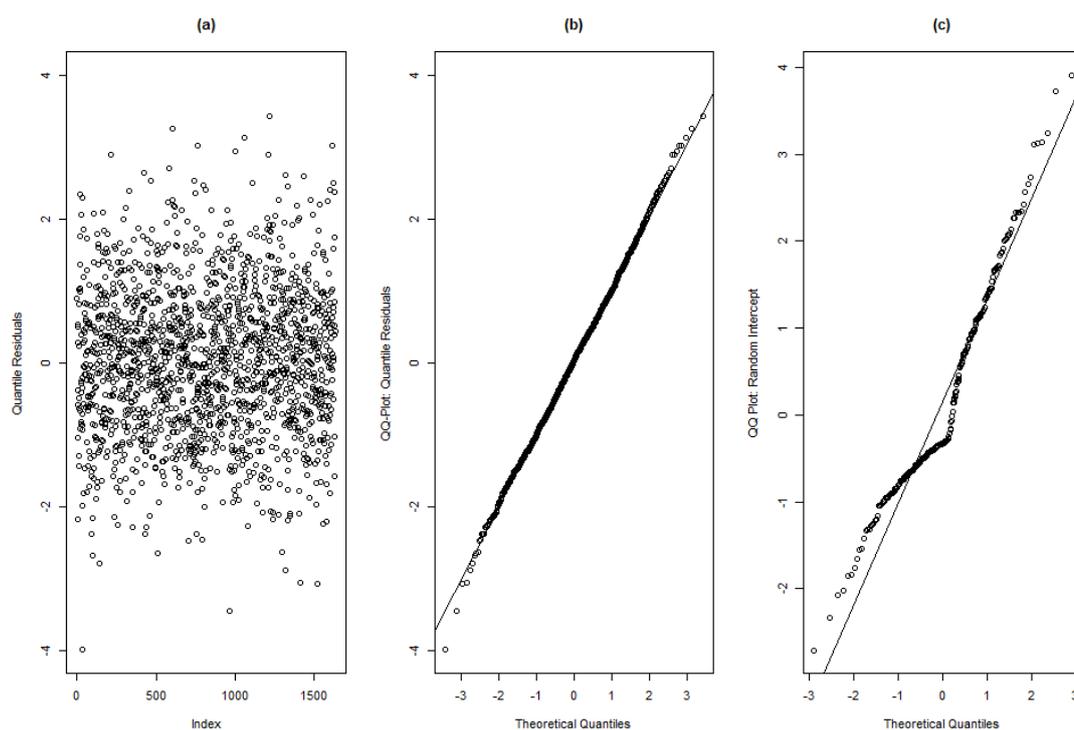
Finalmente, se evalúa en el modelo la relevancia de las actividades de vinculación entre las empresas analizadas y el conjunto de entidades públicas o privadas que forman parte del entorno donde las firmas desarrollan sus actividades productivas. Las actividades de vinculación consideradas incluyen tanto acuerdos cooperativos con participación activa, como otros tipos de vínculos e intercambio informal de información. Surge del modelo que existe relación positiva y significativa entre vinculación y probabilidad de innovar. Esto acuerda con la perspectiva sistémica según la cual la interacción entre firmas y diversas instituciones son factores fundamentales en el éxito del proceso innovador de las empresas (Albornoz *et al.*, 2005). La chance de innovar exitosamente en productos y/o procesos es el triple en aquellas empresas que participaron de actividades de vinculación con otras empresas y/o instituciones comparado con firmas que realizaron sus actividades de innovación de manera individual.

Diagnóstico y capacidad predictiva del modelo

Para realizar el diagnóstico del modelo logístico mixto, siguiendo a Dunn y Smyth (1996), se analiza si los “cuantiles residuales aleatorizados”, siguen una distribución normal. Rigby y Stasinopoulos (2005), utilizan este tipo de residuos para diferentes modelos generalizados aditivos para ubicación, escala y forma (GLAMSS), el cual incluye los modelos generalizados mixtos. En nuestro caso, en el gráfico 4.5 a y b se puede observar que los residuos del modelo se comportan satisfactoriamente. Esto es respaldado por la prueba de normalidad de Shapiro – Wilk la cual conduce a la misma conclusión (p-valor:0,6372 para H_0 : “los residuos se distribuyen normal”).

Por otra parte, uno de los supuestos del modelo es la normalidad de los efectos aleatorios. El gráfico 4.5 c muestra el QQ-plot para la ordenada aleatoria, donde se observa que éste se aleja de la distribución normal. La falta de cumplimiento de este supuesto del modelo es una situación habitual en la práctica. McCulloch y Neuhaus (2011) utilizando estudios de simulación concluyen que, en el modelo logístico mixto, la mayoría de los aspectos de la inferencia estadística son robustos a la mala especificación en la distribución de los efectos aleatorios.

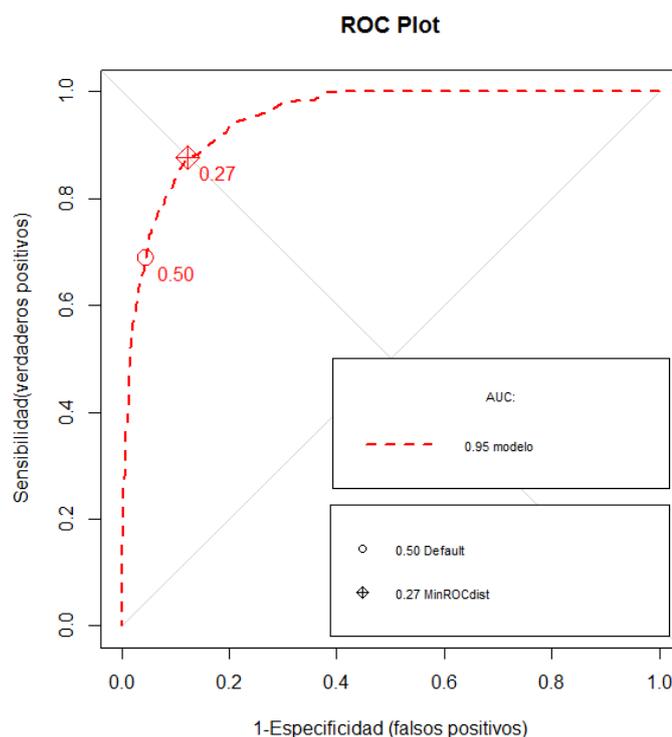
Gráfico 4.5: Diagnóstico del modelo mixto



Para analizar la capacidad predictiva del modelo, tal como fue desarrollado en el punto 3.3.4, se compara la clasificación binaria observada con la clasificación estimada por el modelo y se calculan las medidas de sensibilidad y especificidad. La sensibilidad mide la proporción de empresas innovadoras ($innova=1$) que efectivamente fueron clasificadas como tal por el modelo, en tanto la especificidad mide la proporción de empresas identificadas correctamente como no innovadoras. Son deseables modelos con valores elevados en ambas medidas, pero éstas dependen del punto de corte elegido.

En el gráfico 4.6 se presenta la curva ROC correspondiente al modelo ajustado, la cual muestra todos los pares de sensibilidad y complemento de la especificidad para diferentes puntos de corte. El área bajo la curva (AUC) es de 0,95, la cual muestra una buena capacidad del modelo de discriminar entre empresas innovadoras y quienes no lo son, para cualquier punto de corte.

Gráfico 4.6: Curva ROC



En la tabla 4.7 se presentan las medidas de sensibilidad, especificidad y precisión para un punto de corte de 0,27, el cual fue definido como aquel que minimiza la distancia entre la curva ROC y el punto (0,1). Como puede observarse, el modelo predice correctamente el 87,70% de los casos de empresas que no lograron innovaciones en productos y/o procesos y el 87,61% de los casos de firmas innovadoras.

Tabla 4.7: Capacidad predictiva del modelo

Muestra de entrenamiento	
<i>Especificidad</i>	87,70%
<i>Sensibilidad</i>	87,61%
<i>Precisión</i>	87,68%

Tal cómo surge de los valores de área bajo la curva ROC y las medidas de clasificación expuestas en la tabla 4.7, se obtuvieron muy buenas medidas de la performance del modelo logístico mixto como herramienta predictiva para las empresas de la base de entrenamiento. Incluso si se considera 0,50 como punto de corte para definir la categoría

predicha por el modelo, se alcanzan altos niveles de especificidad y sensibilidad (0,95 y 0,69 respectivamente).

4.3.2 Predicción para nuevos grupos

Para estimar la probabilidad de innovar de aquellas firmas que no formaron parte de la muestra con la que se ajustó el modelo, se consideraron 232 empresas (base de validación), de las cuales 67 (29%) obtuvieron resultados innovadores. Para este conjunto de empresas “nuevas” no se conoce el valor del efecto aleatorio por lo cual es necesario aplicar algún método de predicción de los mismos. Se utilizaron los métodos: “*naive*” basado en el supuesto del modelo mixto sobre el valor esperado de los efectos aleatorios; el “*mejor predictor empírico*” (*EBP*), basado en el predictor empírico Bayesiano; el método de “*regresión lineal*”, centrado en la relación entre las variables independientes y los efectos aleatorios estimados para las unidades de la base de entrenamiento; y el método del “*vecino más cercano*”, basado en la técnica utilizada en clasificación supervisada. Los efectos aleatorios predichos aplicando los distintos métodos se introducen en el predictor lineal de la regresión logística mixta, haciendo posible la clasificación en innovadora o no de la nueva empresa, tal como fue detallado en el punto 3.3.2.

Para la estimación de los efectos aleatorios de nuevas empresas mediante el **método de regresión lineal**, las múltiples variables numéricas disponibles en la base de datos, incluyendo aquellas que no formaron parte del modelo mixto, se agregaron a nivel de empresa considerando la media, la mediana, el mínimo y el máximo. Todas las variables se testearon en la regresión lineal buscando encontrar información adicional que explique el efecto aleatorio. Para seleccionar el modelo se utilizó el procedimiento de *stepAIC* disponible en la librería *MASS* del software R. Este procedimiento utiliza el criterio de información de Akaike (AIC) buscando minimizarlo en cada paso eliminando predictores. El modelo para la ordenada aleatoria seleccionado fue aquel ajustado con las variables

explicativas de los distintos períodos agregadas a nivel de empresa a través de la mediana, y quedó expresado como:

$$\hat{\alpha}_l^* = 0,0642 + 0,0692 AI_EXO_VTA_l$$

Donde $\hat{\alpha}_l^*$ es el valor predicho para la ordenada al origen de la empresa l , la cual forma parte de la muestra de validación, y $AI_EXO_VTA_l$ es la proporción del gasto en actividades de innovación exógenas a la empresa en relación con las ventas.

Al aplicar el **método del vecino más cercano**, las variables numéricas disponibles en la base de entrenamiento para los distintos períodos se agregaron a nivel de empresa utilizando la mediana. Éstas fueron utilizadas para calcular la distancia entre las empresas que contribuyeron a la construcción del modelo mixto (base de entrenamiento) y las nuevas empresas (base de validación). A diferencia de los trabajos de Tamura, *et al.* (2013) y de Caro, *et al.* (2017) en los que se utilizó la distancia Euclídea, la distancia de Minkowski fue la que mejores resultados arrojó. Como medida de centralidad (predicción) de los efectos aleatorios de las empresas de la base de validación, se utilizaron los efectos aleatorios conocidos correspondientes al medoide del grupo de empresas de la base de entrenamiento elegidas como vecinos más cercanos de las empresas nuevas. Donde el medoide es aquel objeto del cluster cuya distancia promedio a todos los objetos que lo componen es mínima.

En la tabla 4.8 se pueden observar las medidas de desempeño del modelo sobre la base de validación según los diferentes métodos aplicados para la predicción de los efectos aleatorios. Se utilizó el valor de 0,27 como punto de corte para el cálculo de las medidas. Este valor corresponde al punto de corte tal que se minimiza la distancia entre la curva ROC y el punto (0,1) para las empresas de la base de entrenamiento. Como se muestra en la tabla, se lograron buenas medidas de la capacidad predictiva del modelo sobre las empresas de la base de validación.

Tabla 4.8: Capacidad predictiva del modelo en empresas de la base de validación

	<i>Métodos de predicción del efecto aleatorio</i>			
	Naive	EBP	Regresión lineal	Vecino más cercano
<i>Especificidad</i>	93,33%	83,03%	92,73%	92,73%
<i>Sensibilidad</i>	38,24%	66,18%	41,18%	44,12%
<i>Precisión</i>	77,25%	78,11%	77,68%	78,54%

Considerando que el método naive supone que los efectos aleatorios de las nuevas empresas son nulos, se procede a comparar los otros tres métodos de estimación.

Al analizar la clasificación correcta de las empresas no innovadoras de la base de validación, se observa que con los métodos de regresión lineal y del vecino más cercano se obtuvieron niveles de especificidad del 92,72%, mientras que con el método EBP se logró clasificar correctamente al 83,03% de las empresas no innovadoras.

El mayor valor de sensibilidad del modelo sobre las empresas de la base de validación se obtuvo utilizando el método de EBP (66,18%). Con los métodos del vecino más cercano y de regresión lineal se logró clasificar correctamente el 44,12% y el 41,18% de las empresas innovadoras respectivamente.

Para cualquiera de los métodos utilizados para estimar los efectos aleatorios el nivel de precisión fue superior al 77%. El nivel de precisión más elevado (78,54%) se obtuvo estimando los efectos aleatorios de las nuevas empresas con el método del vecino más cercano. En segundo lugar, se encuentra el nivel de precisión obtenido con el método EBP (78,11%), seguido por el de regresión lineal (77,68%).

Tal como muestra en Tamura y Giampaoli (2013), aún ante la falta de normalidad de los efectos aleatorios del modelo mixto, tanto el método de EBP como el de regresión lineal mejoran la precisión del modelo. Por su parte, el método del vecino más cercano es el que mejores resultados generó en términos de precisión, resultado que coincide con lo obtenido por diversos trabajos de aplicación (Tamura, *et al.*, 2013; Giampaoli, *et al.*, 2016; Caro, *et al.*, 2017). Al tratarse de una técnica no paramétrica no es afectado por la ausencia de

normalidad de los efectos aleatorios. Todos los métodos utilizados producen buenos resultados, sin diferencias importantes entre ellos.

Capítulo 5: Conclusiones

5.1 Reflexiones finales y contribuciones del trabajo

El ***primer objetivo general*** perseguido en el presente trabajo es el de “*modelar a nivel de firma los principales factores determinantes de la innovación en productos y/o procesos.*”

De esta forma, se contribuye a la comprensión del proceso de innovación, orientando la toma de decisiones tanto en el ámbito empresarial como en el de formulación de políticas públicas.

La innovación es ampliamente reconocida como un mecanismo esencial en la competitividad y el desarrollo económico de largo plazo. A nivel de empresa es un elemento fundamental en la estrategia empresarial para competir en el mercado. La innovación es la herramienta que permite evitar la imitación de productos, acceder a nuevos mercados y hacer uso más eficiente de los recursos generando mejoras genuinas y sostenibles en el tiempo. Es de gran interés identificar las características y acciones a nivel de firma que impactan sobre la probabilidad de obtener resultados innovadores.

Existe una extensa literatura sobre innovación en la que se destacan diversos aspectos de este proceso. Desde la crítica a la economía neoclásica de Schumpeter (1939, 1942) y su definición de innovación como causa motora del desarrollo económico; hacia las ideas evolucionistas y la visión sistémica de la innovación, se genera el marco conceptual para el desarrollo de los estudios empíricos.

La gran cantidad de aspectos involucrados en el proceso de innovación y su importante complejidad lleva a que su estudio pueda ser abordado desde diferentes perspectivas. En el presente trabajo nos centramos en el estudio de la influencia de características estructurales de las empresas, su comportamiento en materia de innovación y su entorno sobre la probabilidad de obtener resultados innovadores.

Se plantea un modelo que permita identificar estas relaciones, utilizando información proveniente de la Encuesta de Innovación y Conducta Tecnológica (EIT), realizada por la Dirección General de Estadística y Censos de la provincia de Córdoba, para los años 2011 a 2016. Es importante destacar el aporte que implica esta investigación en cuanto es el primer análisis sobre esta fuente de datos con representatividad sub - nacional.

Debido a la característica binaria de la variable respuesta (innova / no innova) y dada la estructura de agrupamiento que presentan los datos, reflejado en la falta de independencia entre las observaciones de cada una de las empresas (medidas repetidas), se propone el ajuste de un modelo logístico de ordenada aleatoria. La incorporación al análisis de las mediciones presentadas por cada empresa a través del tiempo constituye un importante aporte de la presente investigación.

El total de empresas disponibles fue separado en dos muestras, una para realizar el ajuste del modelo, compuesta por 276 empresas (muestra de entrenamiento) y otra para realizar inferencia sobre nuevos datos conformada por 232 empresas (muestra de validación). Se delimitó el análisis a la innovación en productos y/o procesos novedosos para el mercado nacional y/o internacional, en pequeñas y medianas empresas (promedio de 5 a 250 empleados en el período analizado).

Respondiendo al **objetivo específico 1**: *“Identificar las principales características de las empresas manufactureras innovadoras en la provincia de Córdoba, para el periodo 2011-2016.”*, se obtienen conclusiones interesantes de las empresas de la muestra.

De la comparación internacional de los indicadores respecto a la proporción de empresas que lograron innovaciones en productos y/o procesos, y los indicadores de los esfuerzos de inversión en actividades de innovación respecto a sus ventas, resulta que las empresas industriales cordobesas se encuentran en niveles similares al total nacional, pero por debajo de los valores registrados en la mayoría de los países miembros de la OECD.

De las empresas bajo estudio surge que, en promedio en el período 2011-2016, el 27% de ellas lograron como resultado de sus esfuerzos innovadores productos y/o procesos nuevos y/o con mejoras significativas novedosos para el mercado nacional o internacional. Se observa que las empresas innovadoras son en promedio de mayor tamaño que las empresas no innovadoras.

En cuanto a los esfuerzos específicos en actividades de innovación, tanto la proporción de empresas que realizaron inversiones, como el porcentaje de sus ventas destinado a estas actividades, son superiores en las empresas que lograron resultados innovadores. Respecto a la composición de los gastos en estas actividades, el más relevante es el destinado a la adquisición de maquinaria y equipo (74%), seguido por la inversión en I+D interna (9%) mostrando una estructura de gasto en actividades de innovación poco equilibrada.

En el grupo de firmas que lograron resultados innovadores exitosos, fue superior la proporción de empresas que mantuvo alguna vinculación con entidades públicas o privadas y como así también la proporción de firmas que utilizaron fondos de programas oficiales de fomento a la innovación para financiar sus inversiones.

Respecto al **objetivo específico 2:** *“Evaluar los principales determinantes de la innovación considerados en la literatura, en el caso de las empresas bajo estudio.”*, los resultados obtenidos en el modelo logístico de ordenada aleatoria ajustado reafirman lo expuesto en el análisis descriptivo y se corresponden con los principales argumentos expuestos en el marco teórico.

En primer lugar, se encontró que a medida que aumenta el tamaño de la firma, mayores son las probabilidades de innovar. Las empresas de mayor tamaño están mejor posicionadas para lograr resultados innovadores en productos y/procesos.

En segundo lugar, las variables referidas a los esfuerzos de las empresas en actividades de innovación, impactan de forma positiva sobre la probabilidad de innovar, en especial

aquellas realizadas dentro de la firma (I+D interna, diseño e ingeniería, capacitación). Adicionalmente, la probabilidad de éxito en el proceso de innovación es superior en las empresas que adoptaron una estrategia basada en la continuidad de la inversión.

Finalmente, surgen como variables relevantes y vinculadas de forma positiva con la probabilidad de obtener resultados innovadores, la utilización de fondos de programas oficiales de fomento a la innovación y la presencia de actividades de vinculación con entidades públicas o privadas, bien sean acuerdos cooperativos con participación activa, como otros tipos de vínculos e intercambio informal de información.

En base a lo expuesto precedentemente se plantean a continuación un conjunto de *“conclusiones que orienten la toma de decisiones empresariales y que permitan el diseño y orientación de políticas públicas en lo referente a innovación.”* (**objetivo específico 4**)

De los resultados obtenidos surge que las pequeñas empresas se encuentran en desventaja frente a las firmas de mayor tamaño. Es por esto importante, que la formulación de políticas se oriente hacia las pequeñas y medianas empresas (PyMes) de manera que dispongan de mejores herramientas para enfrentar los obstáculos y desafíos presentes en el proceso de innovación.

Se mostró que las empresas que realizan mayores esfuerzos en actividades de innovación son aquellas con mayor probabilidad de obtener resultados exitosos, lo cual subraya la necesidad de aplicar políticas tendientes al fomento de la inversión en estas actividades. Políticas tradicionales como el financiamiento público a la I+D, incentivos financieros y fiscales, regulación de los derechos de propiedad industrial e intelectual, ayudan a alcanzar este objetivo. Adicionalmente, para incentivar el mayor desarrollo de actividades de innovación es fundamental el fomento del espíritu emprendedor, el fortalecimiento de capacidades internas en las empresas para innovar y la generación de condiciones institucionales y de mercado que permitan que las mejoras tecnológicas prosperen.

Dada la importancia sobre el éxito innovador de la continuidad de las inversiones en actividades de innovación internas, las políticas públicas deben tender a que no se trate del desarrollo de proyectos esporádicos, sino a que estas actividades sean incorporadas como parte de la rutina de las empresas y de su estrategia de largo plazo.

El modelo pone de manifiesto la importancia de no encarar las acciones de innovación de manera aislada. De lo cual surge la recomendación de profundizar las políticas de apoyo a la articulación institucional, y de estímulo a la generación y fortalecimiento de redes. Es fundamental un entorno institucional que logre coordinar las acciones de las empresas, las instituciones científicas y el Estado, y tal como plantean Gutiérrez Rojas y Baumert (2018) trabajar en la articulación del sistema para poder transformar los avances científicos y desarrollos tecnológicos en productos comercializables.

El **segundo objetivo general** definido para la presente investigación fue el de “*determinar si en el marco de un modelo logístico multinivel, es posible mejorar la predicción de la variable respuesta para unidades que no formaron parte del ajuste del modelo.*”

Se mostró no solo que es posible clasificar a las empresas como innovadoras o no, en función de sus características y comportamiento en materia de innovación, sino también realizar predicción sobre empresas no consideradas en el ajuste del modelo. Guiados por el **objetivo específico 3**, se evaluaron cuatro métodos alternativos para predecir los efectos aleatorios de nuevas unidades: Naive, EBP, Regresión Lineal y Vecino más Cercano. El método del vecino más cercano es el que mejores resultados generó, posiblemente debido a que al tratarse de una técnica no paramétrica no está afectada por la falta de normalidad del efecto aleatorio del modelo mixto.

Si bien las diferencias en el porcentaje de empresas clasificadas correctamente son pequeñas entre un método y otro, la estimación de los efectos aleatorios permitió mejorar la capacidad predictiva del modelo sobre nuevas empresas. De esta manera se contribuye a la estimación de los efectos aleatorios y predicción de la variable respuesta para una nueva observación, en los modelos mixtos.

5.2 Limitaciones del trabajo y futuras líneas de investigación

Para enmarcar las conclusiones expuestas en la sección anterior, es importante mencionar que la principal limitación de este trabajo es que la variable modelada, obtención de resultados innovadores en productos y/o procesos novedosos para el mercado nacional y/o internacional, es definida por la propia firma. Por lo tanto, es un indicador subjetivo que depende de la percepción de la empresa y de sus propios parámetros respecto a lo que constituye una innovación, lo cual incorpora errores de medición de la variable modelada.

Una medida alternativa de innovación utilizada en trabajos referidos principalmente a países desarrollados, es el número de patentes. En el caso de países latinoamericanos y en el caso de Argentina en particular, no es posible tomar este indicador dado el escaso uso de este mecanismo de protección formal de las innovaciones, perdiendo representatividad.

En cuanto a las futuras líneas de investigación que pueden ser planteadas a partir del presente trabajo se destaca la extensión de los métodos de predicción de los efectos aleatorios a otros modelos mixtos. Respecto al estudio del proceso de innovación, el trabajo se ha centrado en los determinantes de la obtención de resultados innovadores. Es posible ampliar este estudio para las empresas industriales de la provincia de Córdoba mediante el análisis de los efectos de la innovación sobre medidas de desempeño de las empresas, como la productividad, competitividad y empleo.

Dos aspectos mencionados en el marco teórico y evaluado en numerosos trabajos empíricos sobre determinantes de la innovación son el sector de actividad de la empresa y su vinculación con mercados externos. Ambas variables fueron evaluadas en el modelo resultando no significativas. Se podría profundizar el análisis de la relación entre innovación y desempeño exportador utilizando modelos que permitan incorporar variables en dos direcciones, como por ejemplo modelos de ecuaciones estructurales.

Referencias Bibliográficas

Albornoz, M., Estébanez, M. E., & Alfaraz, C. (2005). Alcances y limitaciones de la noción de impacto social de la ciencia y la tecnología. *Revista iberoamericana de ciencia tecnología y sociedad*, 2(4), 73-95.

Anlló, G.; Lugones, G. y Peirano, F. (2007). La innovación en la Argentina postdevaluación. Antecedentes previos y tendencias a futuro. En B. Kosacoff (Ed.): *Crisis, recuperación y nuevos dilemas. La economía argentina, 2002-2007*. CEPAL. Disponible en: <http://repositorio.cepal.org/handle/11362/28481>

Arias Ortiz, E.; Crespi, G.; Tacsir, E.; Vargas, F. y Zúñiga, P. (2013). Innovación para el desempeño económico: el caso de las empresas en América Latina y el Caribe. Banco Interamericano de Desarrollo, Nota Técnica 494.

Arza, V., Español, P. y Herrera Bartis, G. (2017). Innovación y exportaciones en la Argentina. Un análisis del comportamiento de las empresas industriales en el marco de un cambio de régimen macroeconómico. *Revista de Economía Política de Bs. As.* Año 11, Vol. 16, 2017, 9-47.

Arza, V., López A. (2010). Innovation and Productivity in the Argentine Manufacturing Sector. IDB Working Paper Series N° 187.

Arza, V.; Español, P.; Herrera Bartis, G. (2017). Innovación y exportaciones en la Argentina. Un análisis del comportamiento de las empresas industriales en el marco de un cambio de régimen macroeconómico. *Revista de Economía Política de Bs. As.* Año 11, Vol. 16, 2017, 9-47.

Astudillo Durán, S. (2018). La innovación en las mipymes manufactureras: estudio comparativo entre el Ecuador y la Argentina. Tesis de doctor en ciencias de la administración. Universidad Nacional del Sur. Bahía Blanca. Argentina.

Bachmann, F. (2016). Los determinantes de la innovación: un aporte para la industria argentina. Tesina de grado. Universidad Nacional de Mar del Plata. Facultad de Cs. Económicas y Sociales.

Barletta, F.; Pereira, M.; Robert, V.; Suarez, D. y Yoguel G. (2013). Innovación y desempeño económico a nivel de firma. Una perspectiva evolucionista. Trabajo N°4 del Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI). Disponible en: <http://www.mincyt.gob.ar/estudios/innovacion-y-desempeno-economico-a-nivel-de-firma-una-perspectiva-evolucionista-9426>

Benavente Hormazábal, J. M. (2005). Investigación y desarrollo, innovación y productividad: un análisis econométrico a nivel de la firma. *Estudios de Economía*, Vol. 32, N° 1, junio, pp. 39-67.

Breslow, N. E. y Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9-25.

Buesa, M.; Baumert, T.; Heijs, J. y Martínez, M. (2002). Los factores determinantes de la innovación: un análisis econométrico sobre las regiones españolas. *Economía industrial*, (347), 67-84.

Callon, M. (1992). The dynamics of techno-economic networks. En: Loombs R., Saviotti P., Alsh V. (ed.), *Technical Change and Company Strategies*. Academy Press, London.

Carlsson, B. y Stankiewicz, R. (1995): On the nature, function and composition of technological systems. En: Carlsson, B. (ed.): *Technological Systems and Economic Performance: The Case of Factory Automation*. Kluwer Academic Publishers. Dordrecht, pp. 21-56.

Caro, N.; Arias, V, Ortiz, P. (2017). Predicción de fracaso en empresas latinoamericanas utilizando el método del vecino más cercano para predecir efectos aleatorios en modelos mixtos. *Revista de métodos cuantitativos para la economía y la empresa* (24). Dic 2017, 5-24 ISSN: 1886-516X. D.L: SE-2927-06. Universidad Pablo Olavide Sevilla España
www.upo.es/revistas/index.php/RevMetCuant/article/view/2878

Chudnovsky, D.; López, A. y Pupato, G. (2004). *Innovation and Productivity in Developing Countries: A study of Argentine Manufacturing Firms' Behaviour (1992-2001)*. Documento de Trabajo 1878, Asociación Argentina de Economía Política, Buenos Aires, 29p.

Cohen, W. y Levinthal, D. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, 99(397), 569-596.

Cooke, P.; Uranga, M. y Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research policy*, 26(4), 475-491.

Crepon, B.; Duguet, E. y Mairesse, J. (1998). Research, Innovation and Productivity: An Econometric Analysis at the Firm Level. *Economics of Innovation and New Technology* 7(2): 115-58.

Crespi, G. y Zúñiga, P. (2010). Innovation and productivity: evidence from six Latin American countries. IDB working paper series N° IDB-WP-218.

Dosi, G. (1997). Opportunities, incentives and the collective patterns of technological change. *Economic Journal* 107, 1530–1547.

Dosi, G. y Nelson, R. (1994). An introduction to evolutionary theories in economics. *Journal of evolutionary economics* 4(3), 153-172.

Dunn, P. y Smyth G. (1996). Randomised quantile residuals. *J. Comput. Graph. Statis.*, 5, 236-244.

Edquist, C. (1997). *Systems of innovation: technologies, institutions, and organizations*. Psychology Press.

Freeman, W. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological cybernetics*, 56(2-3), 139-150.

Giampaoli, V.; Tamura, K.; Caro, P.; Simoes de Araujo, L. (2016). Prediction of financial crisis in Latin American companies using the mixed logistic regression model. *Chilean Journal of Statistics*. Vol. 7, N° 1, April 2016, 31-41.

Gómez, M. C. y Borrastero, C. (2018). Innovación y heterogeneidad productiva en la Industria Argentina. Documentos de trabajo de investigación Facultad de Cs. Económicas. N°1. Abril 2018. Disponible en: <https://revistas.unc.edu.ar/index.php/DTI/>

Griffith, R.; Huergo, E.; Mairesse J. y Peters, B. (2006). Innovation and Productivity across Four European Countries. *Oxford Review of Economic Policy* 22(4): 483-98.

Gutiérrez Rojas, C. y Baumert, T. (2018). Smith, Shumpeter y el estudio de los sistemas de innovación. *Revista de Economía y Política* 5(I), 93-111. Universidad Adolfo Ibañez, Chile. Disponible en: www.economiaypolitica.cl

- Heijs, J. y Buesa, M. (2016). Manual de economía de innovación. Tomo I: Teoría del cambio tecnológico y sistemas nacionales de innovación. Instituto de análisis industrial y financiero. Universidad Complutense de Madrid. ISBN: 978-84-608-5151-6. Disponible en: <https://www.ucm.es/iaif/informacion-sobre-el-instituto-industrial-financiero>
- Hughes, T. (1984). The evolution of large technological systems. En: Bijker W., Hughes T. y Pinch T. (ed.), The Social Construction of Technological Systems. MIT Press, Cambridge.
- Jiang, J. y Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15, 1-96.
- Lee, Y. y Nelder, J. A. (1996). Hierarchical generalized linear models. With discussion. *Journal of the Royal Statistical Society, Series B*, 58:619-678.
- Lee, Y. y Nelder, J. A. (2001). Hierarchical generalized linear models. A synthesis of generalized linear models, random-effect models and structures dispersions. *Biometrika*, 88, 987-10006.
- Levin, R.; Klevorick, A.; Nelson, R.; Winter, S.; Gilbert, R. y Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brookings papers on economic activity*, 1987(3), 783-831.
- Lööf, H.; Heshmati, A.; Asplund, R. y Naas, S.O. (2001). Innovation and Performance in Manufacturing Industries: a Comparison of the Nordic Countries. SSE/EFI Working Paper Series in Economics and Finance No. 457. Agosto 6, 2001
- Lugones, G., Peirano, F., Suárez, D. y Giudicatti, M., (2005). Estrategias Innovativas y Trayectorias empresariales. Documento de Trabajo N° 20, REDES – Centro de Estudios sobre Ciencia, Desarrollo y Educación Superior, Buenos Aires. www.centroredes.org.ar
- Lugones, G.; Suárez, D. y Gregorini, M. (2007). La innovación como fórmula para mejoras competitivas compatibles con incrementos salariales. Evidencia en el caso argentino. Documento de Trabajo N° 36, REDES – Centro de Estudios sobre Ciencia, Desarrollo y Educación Superior, Buenos Aires.
- Lundvall, B. (1992). National systems of innovation: An analytical framework. London: Pinter.
- Lundvall, B. y Johnson, B. (1994). The learning economy. *Journal of Industry Studies* 1 (2), 23–42.
- Mairesse, J. y Mohnen, P. (2010). Using Innovation Surveys for Econometric Analysis. NBER Working Papers 15857. Cambridge, Estados Unidos: Oficina Nacional de Investigación Económica.
- Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, 31(2), 247-264.
- Marín, A.; Liseras, N.; Calá, C. y Graña, F. (2017). Oportunidades de innovación divergentes: ¿es el territorio importante? *Revista Pymes, Innovación y Desarrollo*. Vol. 5, N° 1, 2-23.
- McCullagh, P. y Nelder, J.A. (1989). *Generalized Linear Models*. 2nd Edition, Chapman and Hall, London.
- McCulloch, C., y Neuhaus, J. (2011). Misspecifying the Shape of a Random Effects Distribution: Why Getting it Wrong May Not Matter. *Statistical Science*, Vol 26 N°3, 388-402.
- Milesi, D.; Petelski, N. y Verre, V. (2011). The determinants of innovation: evidence from Argentine manufacturing firms. Universidad Nacional de General Sarmiento. Documento de Trabajo N°3/2011.

- Nelson, R. (1993). National innovation systems: a comparative analysis. Oxford university press.
- Nelson, R. (1995). Recent evolutionary theorizing about economic change. *Journal of economic literature* 33(1), 48-90.
- Nelson, R. y Winter, S. (1982). *An Evolutionary Theory of Economic Change*. The Belknap Press of Harvard University Press. Cambridge.
- OECD (2003), Propuesta de Norma Práctica para Encuestas de Investigación y Desarrollo Experimental de la OECD, "Manual Frascati 2002". Fundación Española para la Ciencia y la Tecnología (FECYT). Publicado por acuerdo con la OCDE, París
- OECD (2005). Oslo Manual - The Measurement of Scientific and Technological Activities. 3rd edition. Guidelines for collecting and interpreting innovation data. Paris, Francia. OECD.
- Pavitt, K.; Robson, M. y Townsend, J. (1989). Technological accumulation, diversification and organization in UK companies, 1945–1983. *Management Science*, 35(1), 81-99.
- Petelski, N.; Milesi, D. y Verre, V. (2017). Financiamiento público a la innovación: impacto sobre esfuerzos tecnológicos en pymes manufactureras argentinas. *Revista Pymes, Innovación y Desarrollo*. Vol.5, N°3, pp 23-44. Disponible en: <https://revistas.unc.edu.ar/index.php/pid/index>
- Rabe-Hesketh S. and Skrondal A. (2012) *Multilevel and Longitudinal Modeling Using Stata*. Third edition. StataCorp LP.
- Raffo, J.; L'huillery, S. y Miotti, L. (2008). Northern and Southern Innovativity: a Comparison across European and Latin American Countries. *European Journal of Development Research*, 20(2), 219-239.
- RICYT, OEA, CYDET (2001) *Normalización de Indicadores de Innovación Tecnológica en América Latina y el Caribe*. Manual de Bogotá. Red Iberoamericana de Indicadores de Ciencia y Tecnología.
- Rigby, R. y Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *Applied. Statistics*, 54 (3), 507–554.
- Rosenberg, N. (1974). Science, invention and economic growth. *The Economic Journal*, 84(333), 90-108.
- Rothwell, R. (1994). Towards the fifth-generation innovation process. *International Marketing Review*, vol. 11, n°1. Pp 7-31.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78:719-727.
- Schmookler, J. (1962). Economic sources of inventive activity. *The Journal of Economic History*, 22(01), 1-20.
- Schumpeter, J. (1939). *Business cycles*. New York: McGraw-Hill.
- Schumpeter, J. (1942). *Capitalism, socialism and democracy*, New York, Harper.
- Skrondal, A., Rabe-Hesketh, S., (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC. ISBN 1-58488-000-7.

Tamura, K. A. y Giampaoli, V. (2010). Prediction in multilevel logistic regression. *Communications in Statistics – Simulation and computation*, 39 (6), 1063-1076.

Tamura, K.A. y Giampaoli, V. (2011). Prediction for an observation in a new cluster for mixed Logistic Regression considering k random coefficients. In: 26th International Workshop on Statistical Modelling. Vol. 1, pp. 593-596.

Tamura, K.A. y Giampaoli, V. (2013). New prediction method for the mixed logistic model applied in a marketing problem. *Computational Statistics & Data Analysis*, v. 66, 202-216, 201.

Tamura, K.A., Giampaoli, V. y Noma, A. (2013). Nearest Neighbors Prediction Method for mixed logistic regression. In: 28th International Workshop on Statistical Modeling, Palermo, 799-802.

Velasco Balmaseda, E.; Zamanillo Elguezabal, I. y Gurutze Intxaurburu, C. (2007). Evolución de los modelos sobre el proceso de innovación: Desde el modelo lineal hasta los sistemas de innovación. XX Congreso anual de AEDEM, Vol. 2, 2007. Disponible en: <http://dialnet.unirioja.es/servlet/articulo?codigo=2499438>

Zweig, M. y Campbell, G. (1993). Receiver operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39:561–577.

Anexos

A1 Cuestionario Encuesta de Innovación y Conducta Tecnológica Córdoba



#N/A	#N/A
EIT ENCUESTA SOBRE INNOVACION Y CONDUCTA TECNOLOGICA CORDOBA AÑO DE REFERENCIA 2016 CARÁCTER ESTRICTAMENTE CONFIDENCIAL Y RESERVADO - Ley Nac. Nº 17.622 - Ley Prov. Nº 5.454 (Los datos serán publicados solamente en compilaciones de conjunto para no violar el secreto comercial Art. 10 de la Ley Nac. Nº 17.622 y Art. 5 de la Ley Prov. Nº 5.454)	
Esta encuesta tiene como objetivo principal obtener la información necesaria para calcular indicadores de Innovación e Investigación y Desarrollo (I+D) y sobre el uso de las Tecnologías de la Información y de las Comunicaciones (TICs) correspondientes al año de referencia.	
UNIDAD DE RELEVAMIENTO Y PERIODO DE REFERENCIA Los datos solicitados corresponden a empresas asentadas en la provincia de Córdoba y al año de referencia	
FORMA DE ANOTAR LOS DATOS Debe marcar con una cruz la/s casilla/s correcta/s. Todos los datos referentes a valores monetarios se solicitan en pesos, omitiendo los centavos.	
PLAZO DE RESPUESTA El presente cuestionario, de conformidad con lo establecido por la Ley Nac. Nº 17.622 y Ley Prov. Nº 5.454, deberá ser devuelto cumplimentado por correo electrónico como archivo adjunto a la dirección de correo electrónico eit@cba.gov.ar , perteneciente a la Dirección General de Estadística y Censos sito en Ituzalngó 1351 de la ciudad de Córdoba, en un plazo no superior a los 15 días posteriores a su recepción.	
CONSULTAS Si desea hacer alguna consulta puede dirigirse a: Dirección General de Estadística y Censos (División Encuestas a Empresas), Ituzalngó 1351 - Nueva Córdoba (X5000KKA), por mail a: eit@cba.gov.ar o por teléfono al 0351-4476094	
OBLIGATORIEDAD DE RESPONDER Ley Nac. 17.622 - Artículo 11: Todos los organismos y reparticiones nacionales, provinciales y municipales, las personas de existencia visible o ideal, públicas o privadas con asiento en el país, están obligados a suministrar a los organismos que integran el Sistema Estadístico Nacional, los datos e informaciones de interés estadístico que éstos le soliciten. Ley Nac. 17.622 - Artículo 15: Incurrirán en infracción y serán pasibles de multas conforme al procedimiento que se establezca en la reglamentación de la presente ley, quienes no suministren en término, falseen o produzcan con omisión maliciosa las informaciones necesarias para las estadísticas y los censos a cargo del Sistema Estadístico Nacional. Ley Prov. 5.454 - Artículo 4: Todas las reparticiones e instituciones dependientes de los poderes públicos de la provincia y municipalidades y las personas de existencia visible o ideal, públicas o privadas están obligadas a proporcionar a la Dirección de Informática, Estadística y Censos o a los organismos integrantes del Servicio Estadístico Provincial, todo dato e informe de interés estadístico que le soliciten y dentro de los plazos fijados a tales fines. Ley Prov. 5.454 - Artículo 2: Incurrirán en infracción y serán pasibles de multas, quienes no suministren en término, falseen o produzcan con omisión maliciosa las informaciones y censos requeridos por la Dirección General de Estadística y Censos dependiente de la Secretaría General de la Gobernación de la Provincia de Córdoba.	
CONFIDENCIALIDAD DE LOS DATOS SUMINISTRADOS Ley Nac. 17.622 - Artículo 10: Las informaciones que se suministran a los organismos que integran el Sistema Estadístico Nacional, en cumplimiento de la presente ley serán estrictamente secretas y sólo se utilizarán con fines estadísticos. Los datos deberán ser suministrados y publicados en compilaciones de conjunto, de modo que no pueda ser violado el secreto comercial o patrimonial, ni individualizarse las personas o entidades a quienes se refieren. Quedan exceptuados del secreto estadístico los siguientes datos de registro: nombre y apellido o razón social, domicilio y rama de actividad. Ley Prov. 5.454 - Artículo 5: Las informaciones que se suministran en cumplimiento de la presente ley, serán estrictamente secretas y sólo se utilizarán con fines estadísticos. Los datos así receptados, deberán ser proporcionados y publicados, exclusivamente en compilaciones de conjunto, de modo que no pueda ser violado el secreto comercial, patrimonial o profesional, ni individualizarse las personas o entidades a quienes se refieren. Quedan exceptuados del secreto estadístico los siguientes datos de registro: nombre y apellido o razón social, domicilio y rama de actividad.	

En Paso de los Andes

Shirley 17 0

DATOS DE LA EMPRESA al 31/12/2016

Nombre	#N/A
Domicilio de la Administración Central	#N/A
Entre calle (calle1)	#N/A
y calle (calle2)	#N/A
Código Postal	#N/A
Localidad	#N/A
Provincia	#N/A

Principal actividad industrial	#N/A	01		
Nº de CUIT	#N/A	02		
Página Web	#N/A	03		
Forma jurídica	#N/A	#N/A		
Número de locales industriales que posee la empresa actualmente en Córdoba y en el país (Local industrial es todo espacio físico aislado o separado de otros, utilizado para desarrollar actividades industriales).	País	#N/A	17	
	Córdoba	#N/A	038	
La empresa es Independiente o parte de un grupo:	#N/A	#N/A	#N/A	051
Año de inicio de las actividades	#N/A	302		
Porcentaje de capital extranjero en el capital total de la empresa	#N/A	%	20	
Para quienes hayan declarado en el punto anterior tener capital extranjero en el capital total de la empresa, ¿Cuál es el país de origen del mismo?	#N/A	301		

by Carlo & Jofre



DESEMPEÑO ECONÓMICO EN EL AÑO 2016

1. VENTAS		
Ventas de la empresa (comprende la venta realizada por la empresa a precio de venta en fábrica, de bienes elaborados en su local o mandados a elaborar a terceros con materia prima de la empresa y la reventa de productos). No debe incluir IVA, ni impuestos internos, ni a los combustibles, tampoco reintegros ni subsidios. En la venta de bienes de capital producidos por la empresa debe incluir el ingreso por la instalación cuando el bien se instale por cuenta del productor. (monto en pesos, omitiendo centavos)		
1.1 ... bienes producidos por la empresa al Mercado Interno		230
1.2 ... bienes producidos por la empresa al Mercado Externo		251
1.3 Reventa de productos terminados (comprende la venta realizada por la empresa de mercaderías adquiridas a otras empresas. No debe incluir IVA, ni impuestos internos, ni a los combustibles).		252
Ventas Totales		253
1.4 Del total de ventas informado en este ítem, qué monto corresponde a los locales asentados en la Provincia de Córdoba		206

2. RECURSOS HUMANOS SEGÚN SU EDUCACIÓN Y FUNCIÓN EN LA EMPRESA (al 31 de diciembre) Se requiere la cantidad de personal de cada nivel de educación formal alcanzado y que se desarrolló en su rutina normal de trabajo en las siguientes funciones en la empresa.	CANTIDAD DE PERSONAL					
	Gerentes	Administrativos de sistemas	Otros administrativos	Operarios	Otras funciones	Total
2.1 Profesionales de Ingeniería o de Otras Ciencias Duras (formación universitaria de grado en alguna Ingeniería (Electrónica, Sistemas, Civil, Industrial, etc.) u otra ciencia dura (Química, Física, Matemática, Biología, etc.).	755	815	900	825	830	86
2.2 Otros Profesionales (aquellos que hayan completado su formación universitaria de grado y que no hayan sido asentados en 2.1)	756	816	901	826	831	87
2.3 Personal con Educación Técnica (aquellos con secundario completo o terciario)	757	817	902	827	832	85
2.4 Personal con Educación Básica o Inferior (aquellos que NO han completado la escuela secundaria [sea Bachillerato, Técnico o Comercial] o el polimodal).	758	818	903	828	833	84
Total RR. HH.	759	819	904	829	834	88
2.5 De este total de RR HH, qué cantidad corresponde a la Provincia de Córdoba						807

by Cueto & Jofre

ACTIVIDADES DE INNOVACIÓN EN EL AÑO 2016

La pregunta 3 es una de las más importantes del cuestionario. No se apresure para completarla, al contrario, dedíquese el tiempo que sea necesario ya que el esfuerzo que realice en este momento le facilitará muchísimo la tarea con el resto de las preguntas. Primero intente identificar cuáles de las siguientes Actividades de Innovación ha desarrollado la empresa durante el período.

En el caso de que haya realizado algunas de las actividades mencionadas anteriormente recuerde que debe cuantificar los recursos destinados a las mismas en la columna correspondiente. Se solicitan los montos en pesos, netos de impuestos, de los gastos realizados en el año. Si bien pueden no ser exactos, deberán reflejar los recursos destinados a la realización de cada una de las actividades de innovación. En caso de que se dificulte su cuantificación, no deje el casillero vacío, complete con algún valor aproximado. Si solo puede cuantificar el gasto total, aproxime el gasto de cada actividad que haya realizado en el período con el porcentaje que corresponde del total.

Luego se pasará a completar la columna de la tabla que se refiere a los Resultados (cambios, mejoras y/o innovaciones) obtenidos de las actividades de innovación desarrolladas en el período. Si el resultado de una determinada actividad de innovación ha sido positivo, es decir alcanzó algún cambio, mejoras y/o innovaciones de proceso, producto o técnica organizacional o de comercialización consigne la opción Sí obtuvo. En cambio, si no obtuvo resultados positivos deberá distinguir si es que la actividad de innovación aún no ha madurado (Aún no obtuvo resultados positivos, pero sigue haciendo la actividad) o si es que de la actividad arrojó resultados negativos y ha dejado de desarrollar la actividad (No obtuvo resultados y Abandonó la actividad).

ACTIVIDADES DE INNOVACIÓN DESARROLLADAS

3. Indique qué actividades de innovación realizó en procura de lograr innovaciones de producto, de procesos, de tipo organizacional o de comercialización, y cuantifique los recursos destinados a cada actividad (Marque con una X la opción que corresponda)	Sí		No		(Monto en pesos sin centavos)	Obtuvo resultados?		
						No	Aún no	Sí
3.1 I+D interna				X	200			200
3.2 I+D externa				X	281			281
3.3 Adquisición de Maquinarias y equipos				X	282			282
3.4 Adquisición de Hardware				X	283			283
3.5 Adquisición de Software				X	284			284
3.6 Contratación de Tecnología				X	285			285
3.7 Capacitación				X	286			286
3.8 Diseño industrial y actividades de ingeniería				X	287			287
3.9 Consultorías				X	288			288
TOTAL					289			289

4. ¿Cuál de las siguientes opciones representa el inicio del proceso innovativo de la empresa?		
4.1 Detección de una demanda total o parcial insatisfecha en el mercado		205
4.2 Aprovechamiento de una idea o de novedades científicas y técnicas		206
4.3 Otros, especificar:		207

5. ¿Cuáles de los siguientes objetivos orientaron las Actividades de Innovación?		
5.1 Uso de nuevos materiales		260
5.2 Reorganización del proceso productivo		261
5.3 Incorporación de nuevos bienes de capital		262
5.4 Cambios en los canales de distribución		263
5.5 Cambios en la estrategia de marketing		264
5.6 Mejora en el aprovechamiento de la mano de obra		265
5.7 Modificación en las prestaciones o características del producto		266
5.8 Lanzamiento de nuevos productos		267
5.9 Búsqueda de nuevos nichos o mercados internos		268
5.10 Búsqueda de nuevos nichos o mercados externos		269
5.11 Reducir costos		236
5.12 Otros objetivos, especificar:		237

6. Si "NO" realizó actividades de I+D interna, por favor indique los motivos:	No importante	Poco importante	Moderadamente importante	Muy importante	
6.1 El tamaño pequeño del mercado no permite recuperar la inversión en I+D					306
6.2 La inversión en I+D es muy riesgosa					307
6.3 La inversión en I+D es muy costosa					308
6.4 Falta de acceso al crédito					309
6.5 Dificultades para apropiarse de los resultados de I+D					310
6.6 Falta de apoyo del sector público					311
6.7 Las fuentes externas proveen información suficientes para la I+D					312
6.8 Las universidades proveen el conocimiento que se obtendría haciendo I+D					313
6.9 Los institutos públicos proveen el conocimiento que se obtendría haciendo I+D					314
6.10 La actividad desarrollada por la empresa no requiere I+D					600
6.11 La empresa no está interesada en realizar actividades de I+D					601
6.12 La empresa terciariza la I+D interna desarrollando actividades de I+D externa					602
6.13 Las actividades de I+D interna las realiza la casa matriz, otras empresas del grupo, etc					603
6.14 Falta de RRRH capacitados/falta de capacidad financiera para contratarlos					905
6.15 Otras Razones, especificar:					004

FUENTES DE FINANCIAMIENTO DE LAS ACTIVIDADES DE INNOVACIÓN E INVESTIGACIÓN Y DESARROLLO INTERNA

En la pregunta 7, resulta de especial interés conocer de dónde provienen los fondos utilizados para solventar las actividades de innovación (presentadas en la tabla de la pregunta 3). Si la empresa ha realizado I+D, adquirido bienes de capital, modificado el soporte informático o replanteado sus métodos de gestión o comercialización con la intención de lograr mejoras o adelantos, se buscará saber cómo y de dónde ha obtenido los fondos requeridos para llevar adelante dichas actividades de innovación. Si bien la pregunta requiere de una respuesta numérica, no es necesario arribar a cifras extremadamente precisas, admitiéndose como válidos valores orientativos o aproximados.

7. FUENTES DE FINANCIAMIENTO	I+D interna		Innovación (incluye I+D interna)	
	El total de esta columna deberá ser igual al gasto en I+D interna del cuadro 3		El total de esta columna deberá ser igual al gasto total de las Actividades de Innovación del cuadro 3	
	Monto en pesos de unidades		Monto en pesos de unidades	
7.1 Recursos Propios		906		907
7.2 Recursos de otra procedencia:				
7.2.1 Provenientes de la banca comercial pública o privada		65		80
7.2.2 Provenientes de proveedores		58		79
7.2.3 Provenientes de clientes		59		74
7.2.4 Provenientes de otras Empresas		60		75
7.2.5 Provenientes de organismos públicos y/o privados		61		76
7.2.6 Provenientes de otros organismos públicos de fomento		62		77
7.2.7 Provenientes de fundaciones, asociaciones sin fines de lucro y Organismos No Gubernamentales		64		78
7.2.8 Provenientes de universidades públicas		888		880
7.2.9 Provenientes de universidades privadas		339		881
7.2.10 Provenientes de organismos internacionales (Banco Interamericano de Desarrollo, Banco Mundial, Unión Europea, etc.)		66		81
7.2.11 Provenientes de otras fuentes		67		82
TOTAL		98		83

RELACIONES CON LOS PROGRAMAS DE FINANCIAMIENTO PARA LAS ACTIVIDADES DE INNOVACIÓN

8. Indique si tiene conocimiento y ha utilizado fondos de las siguientes instituciones y programas oficiales de estímulo a las actividades de innovación	¿Tiene conocimiento?		¿Ha utilizado fondos?		
	Si	No	Si	No	
8.1 FONAR			626		629
8.2 FONCYT			627		630
8.3 FORSOFT			710		711
8.4 FONTEC			908		914
8.5 PRODIS			909		915
8.6 FONDIS			910		916
8.7 Innovación en Seguridad Ciudadana (I+5)			911		917
8.8 Innovación en Accesibilidad (I+A)			912		918
8.9 Otros programas o instrumentos del Sector Público (especificar)			913		919

¿Cuáles han sido las principales razones para no utilizarlos?	
a. Proyectos rechazados	632
b. Tasas de interés elevadas	633
c. Excesiva exigencia de garantías	634
d. Dificultades burocráticas	635
e. Dificultad para formular proyectos de innovación	636
f. Otros (especificar):	637

PERSONAS DEDICADAS A ACTIVIDADES DE INNOVACIÓN

9. Se solicita que detalle la cantidad de empleados que realizaron actividades de innovación, particularmente de I+D interna y Diseño Industrial e Ingeniería, ya sea en unidades o departamentos "Formales" (unidades específicamente abocadas a esas actividades) o de manera "NO Formal" (dentro de la empresa, pero no en unidades específicas, sino de manera simultánea o conjunta con otras actividades). En esta pregunta, es irrelevante si la dedicación es a tiempo completo o parcial, ya que solo debe contabilizarse la cantidad de personal.	Cantidad de personal	
	Unidad formal	Unidad NO formal
9.1 Investigación y desarrollo (I+D)	94	97
9.2 Diseño Industrial y actividades de Ingeniería	95	98



Indique cuántas de las personas que trabajan en una <u>unidad formal o no formal de I+D interna</u> poseen los siguientes títulos.		
a. Título terciario técnico		835
b. Título de Grado		303
c. Título de Maestría		304
d. Título de Doctorado		708

10. Desglose el personal de I+D formal (variable 94, en la pregunta 9.1) según su función en ese departamento ACLARACIÓN: Para los fines de esta encuesta, se considera "investigador" a cualquier persona que gestiona o dirige el grupo de trabajo dedicado a I+D, aunque la misma no sea reconocida en la empresa con el título de investigador. Puede suceder que una persona considerada como "técnico" esté ocupando el puesto de investigador ya que cumple tareas de directivo o encargado de la I+D. Por tales motivos, debe haber al menos 1 (una) persona en carácter de "investigador".	Cantidad total de personas	Cantidad de Mujeres
10.1 Investigadores y Asistentes de Investigación: Comprende los profesionales que trabajan en la creación de nuevos conocimientos, productos, procesos, métodos y sistemas y en la gestión de los respectivos proyectos, incluyendo sus directores o administradores de proyectos. También se incluye a los profesionales que realizan actividades de investigación y desarrollo bajo la dirección de un investigador, con la finalidad de formarse y que por ello reciben una remuneración.	100	709
10.2 Técnicos: Son personas que ejecutan sus tareas bajo la supervisión de un investigador. En general comprende a: asistentes de laboratorio, dibujantes, asistentes de ingeniero, fotógrafos, técnicos mecánicos y eléctricos, programadores, etc.	102	782
10.3 Otros: Son personas que colaboran con servicios de apoyo a las actividades de investigación y desarrollo, tales como personal de oficina, operarios, etc. Esta categoría incluye a gerentes y administradores, que se ocupan de problemas financieros, de personal, etc. siempre que sus actividades se relacionen totalmente o en parte con la investigación y desarrollo.	103	763
Total	104	764



GASTOS DE INVESTIGACIÓN Y DESARROLLO INTERNA

11. Distribuya el gasto en I+D interna según su tipo RECORDAR: el total declarado en las columnas del cuadro (esto es: la suma de las variables 49 y 53), deberá ser igual al valor declarado en las variables 280 del cuadro 3 (gasto en I+D interna). No es necesario un valor exacto, admitiéndose como válidos valores orientativos.	Gastos en personal (asalariado o de cualquier otro tipo) Monto en pesos sin centavos	Otros gastos (corrientes y de capital) Monto en pesos sin centavos
11.1 Investigación Fundamental o Básica: Consiste en trabajos principalmente teóricos que se emprenden para obtener nuevos conocimientos, sin prever en darles ninguna aplicación determinada. Es el caso, por ejemplo, de los trabajos investigativos desarrollados en los centros de investigación universitarios.	46	50
11.2 Investigación Aplicada: Consiste en trabajos realizados para adquirir nuevos conocimientos pero fundamentalmente, dirigidos hacia un objetivo práctico específico.	47	51
11.3 Desarrollo Experimental: Consiste en trabajos sistemáticos basados en los conocimientos existentes derivados de la investigación y/o la experiencia práctica, dirigidos a la producción de nuevos materiales, productos y dispositivos; al establecimiento de nuevos procesos, sistemas y servicios, o a la mejora sustancial de los ya existentes.	48	52
Total	49	53



RELACIONES CON EL SISTEMA NACIONAL DE INNOVACION

Pregunta 12: La innovación tiene un fuerte carácter sistémico, es decir, las empresas desarrollan actividades de innovación que están estimuladas, desalentadas o influidas por su entorno o ámbito de acción. En este sentido, las empresas no innovan de forma aislada sino que se nutren de (y nutren a) otras empresas e instituciones.

Se recomienda, en primer lugar, identificar las "Instituciones" o agentes y definir si "tuvo alguna vinculación". Cabe aclarar que por "vinculación" debe entenderse cualquier tipo de relación, ya sea formal (un contrato, un convenio, un joint venture) o informal o que haya involucrado o no una retribución monetaria de alguna de las partes. Tampoco deben restringirse las relaciones al ámbito nacional: si una empresa ha intercambiado información con una empresa argentina, debe ser considerado como una relación con "otras empresas" con el objeto de intercambiar u obtener "información".

12. ¿Tuvo vinculación con las INSTITUCIONES del Sistema Nacional de Innovación? (Si NO mantuvo vinculación pase a la pregunta 14)	¿Con qué objetivos?					Tipo de vinculación		
	Intercambio de información	Capacitación	Asesoría y apoyo técnico	Ingeniería y Diseño Industrial	I+D	Acuerdos cooperación con participación activa	Otros vínculos / intercambio informal de información	
12.1 Universidades públicas						921		942
12.2 Centros de Formación						922		938
12.3 INTI (Instituto Nacional de Tecnología Industrial)						923		935
12.4 INTA (Instituto Nacional de Tecnología Agropecuaria)						924		936
12.5 CONICET						925		944
12.6 Agencia Nacional de Promoción Científica y Tecnológica						926		937
12.7 Otros organismos o programas gubernamentales de ciencia y tecnología						927		938
12.8 Universidades privadas						928		943
12.9 Casa Matriz / Otras empresas del grupo						929		939
12.10 Otras empresas no relacionadas						930		940
12.11 Proveedores						931		941
12.12 Clientes						932		942
12.13 Laboratorios / Empresas de I+D						933		943
12.14 Centros / consultores Tecnológicos						934		944

13. En caso de haber mantenido vinculaciones con universidades e institutos públicos de investigación (opciones 12.1 a 12.7) y no estar satisfecho con los resultados, ¿cuáles fueron las causas de su insatisfacción?. Indique el nivel de importancia.	No importante	Poco importante	Moderadamente importante	Muy importante	
13.1 Diferencias en los tiempos					639
13.2 Diferencias en puntos de vista y/u objetivos					640
13.3 Los investigadores de la universidad/institutos están demasiado orientados a la ciencia básica					641
13.4 Los investigadores de la universidad/institutos no están lo suficientemente orientados a la ciencia básica					642
13.5 Baja respuesta de las universidades/institutos a las demandas de la firma					643
13.6 Diferencias con respecto a la apropiabilidad de los resultados del proyecto (problemas de propiedad intelectual)					644
13.7 Falta de preparación del personal de la firma para tratar con la universidad/institutos					645
13.8. Otra (especifique):					646

14. En caso de no haber mantenido vinculaciones con universidades e institutos públicos de investigación (opciones 12.1 a 12.7) indicar el por qué. Califique las siguientes razones en correspondencia con sus niveles de importancia.	No importante	Poco importante	Moderadamente importante	Muy importante	
14.1 La I+D de la firma es suficiente fuente de conocimiento					647
14.2 Las universidades/institutos no comprenden nuestra línea de negocios					648
14.3 Los acuerdos contractuales son difíciles					649
14.4 Falta de confianza					650
14.5 Baja calidad de la investigación					651
14.6 Las universidades se preocupan solamente por la ciencia básica					652
14.7 Distancia geográfica					653
14.8 Dificultades en el diálogo por ritmos diferentes					654
14.9 Cuestiones de propiedad intelectual					655
14.10 Falta de información sobre las actividades desarrolladas por las instituciones					656
14.11 Otra (especifique)					686

* Debe contestar esta pregunta

by C&C & J&G



INNOVACIONES LOGRADAS Y MECANISMOS DE PROTECCIÓN EN EL AÑO 2016

Definidas las actividades de innovación realizadas intente identificar qué **ORIENTACIÓN** pretendía la empresa mejorar con el desarrollo de cada una de ellas: **Productos Nuevos o Mejorados**: (todo lo relacionado con modificaciones a las características de los productos ya existentes o al surgimiento de nuevos productos), **Procesos Nuevos o Mejorados**: (todo lo relacionado a cómo se produce), **Organización**: (todo lo relacionado a cómo se organiza y articula el trabajo de las distintas áreas de la empresa) y **Comercialización**: (todo lo relacionado cómo la empresa se vincula con sus clientes o con el mercado – distribución, modalidades de venta–).

15. ¿Las innovaciones logradas fueron novedosas para la empresa y/o el mercado nacional y/o el mercado internacional?	La empresa		El mercado nacional		El mercado internacional		
	SI	NO	SI	NO	SI	NO	
15.1 Nuevo producto			315		321		571
15.2 Mejoró significativamente un producto existente			316		322		572
15.3 Nuevo proceso			317		323		573
15.4 Mejoró significativamente un proceso existente			318		324		574
15.5 Innovaciones organizativas			319		325		575
15.6 Innovaciones en comercialización			320		326		576

16. ¿Qué incidencia tuvieron las innovaciones logradas sobre las ventas?	No tuvo incidencia	Sí tuvo incidencia				
		Baja	Media	Alta	Muy Alta	
16.1 Innovación de producto						845
16.2 Innovación de proceso						846
16.3 Innovación organizacional						847
16.4 Innovación de comercialización						848

17. Distribuya porcentualmente el valor de las ventas totales según el grado de novedad de la innovación de producto alcanzada en cada caso.	Ventas Totales (%)
17.1 Nuevos o significativamente mejorados para la empresa y para el mercado (local y/o internacional)	206
17.2 Nuevos o significativamente mejorados para la empresa, pero ya existentes en el mercado	207
17.3 Iguales o que no fueron alterados significativamente de los anteriores de la empresa	208
TOTAL	%

18. Indique la cantidad de patentes solicitadas y obtenidas en Argentina, Estados Unidos, Europa y el resto del mundo.	Argentina	Estados Unidos	Europa	Resto del mundo
18.1 Patentes solicitadas	618	620	622	624
18.2 Patentes obtenidas	619	621	623	625

19. Indique cuales han sido los mecanismos de protección de los productos nuevos o significativamente mejorados llevados adelante por la empresa.			
<u>Métodos de protección formal</u>		<u>Métodos de protección estratégica</u>	
a) Marca	608	f) Controlar las redes de distribución	613
b) Patentes	609	g) Llegar primero al mercado	614
c) Diseño Industrial	610	h) Economía de escala	615
d) Denominación de origen	611	i) Secreto Comercial	616
e) Derechos de autor	612	j) Complejidad del diseño	617

24. ¿Cómo se distribuyeron los montos invertidos en TICs para los siguientes bienes y servicios? ACLARACIÓN: El total de este cuadro deberá contener un valor igual o mayor al total registrado en el total del cuadro 23, variable 673.		
24.1 Desarrollo de Software o sistema específico		674
24.2 Adaptación o introducción de un Software específico		675
24.3 Instalación de Hardware y equipos de comunicación		676
24.4 Otras Inversiones en TICs		677
24.5 Desarrollo de página web		714
TOTAL:		678

25. Las inversiones en TICs, incluyeron:	SI	NO	
25.1 Desarrollo de software o sistema específico			115
25.2 Adaptación de un software estándar (paquete)			116

26. Las actividades de la pregunta 25 ¿a cargo de quién estuvieron? (múltiples respuestas posibles)		
26.1 Personal propio de la empresa		117
26.2 Consultora		118
26.3 Proveedor		119
26.4 Otros (especifique):		120

27. ¿Cuál fue el origen de los fondos utilizados para solventar las inversiones en TICs? Si la empresa ha realizado inversiones en alguno de los ítems anteriores, se buscará saber cómo y de dónde ha obtenido los fondos requeridos para llevar adelante dichas acciones. Indique si la empresa ha destinado recursos propios y/o de otra procedencia. Si bien la pregunta requiere de una respuesta numérica, no es necesario arribar a cifras extremadamente precisas, admitiéndose como válidos valores orientativos o aproximados. Es importante completar con cero (0) los casilleros de las categorías vacías y comprobar que la suma sea el 100% de los fondos utilizados en TICs. Aclaración: El total de este cuadro deberá ser igual al total del cuadro 24 e igual o mayor al total registrado en el cuadro 23	TICs Monto en pesos sin centavos
27.1 Recursos Propios	935
27.2 Recursos de otra procedencia:	
27.2.1 Provenientes de la banca comercial pública o privada, de organismos internacionales (Banco Interamericano de Desarrollo, Banco Mundial, Unión Europea, etc.)	702
27.2.2 Provenientes de proveedores, clientes, de otras empresas	703
27.2.3 Provenientes de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT): Fondo para la Inversión Científica y Tecnológica (FONCYT); Proveniente del Fondo Tecnológico Argentino (FONTAR); y/o Fondo Fiduciario de Promoción de la Industria del Software (FONSOFIT)	855
27.2.4 Provenientes de otros organismos públicos de fomento	849
27.2.5 Provenientes de universidades (públicas o privadas)	850
27.2.6 Provenientes de fundaciones, asociaciones sin fines de lucro y Organismos No Gubernamentales	705
27.2.7 Provenientes de otras fuentes	706
TOTAL	707

28. ¿Cuántas computadoras tenía la empresa?		126	✘
---	--	-----	---

29. ¿Cuál fue el número total de personas que utilizó PC en su rutina normal de trabajo?		128	✘
--	--	-----	---

30. ¿Cuántos servidores tenía la empresa?			
Cantidad de servidores		712	✘
No tiene		713	

31. ¿La empresa utilizaba Internet?			
Sí		129	✘
No			

32. ¿Cuál fue el número total de personas que utilizaba Internet, en su rutina normal de trabajo? (Se consideran tanto los empleados que utilizan computadora en su rutina normal de trabajo como quienes no utilizan computadora)		131	
--	--	-----	--

33. ¿La empresa contaba con un sitio Web propio? ¿O participaba en un sitio Web donde haya tenido control sobre el contenido?			
Sí			132
En construcción			
No			

34. ¿La empresa recibió ordenes de compra a través de Internet?			
Sí		137	
No			

35. ¿La empresa ha vendido por internet?			
Sí		683	
No			

36. Para las empresas que hayan vendido por internet, ¿cuáles de los siguientes beneficios consiguió a través de las ventas por internet? (respuestas múltiples posibles)		
36.1 Reducción del tiempo de transacción		687
36.2 Mayor calidad de servicios para el consumidor		688
36.3 Menor costo de negocio		689
36.4 Mayor volumen de ventas y/o número de consumidores		690
36.5 Equipararse con la competencia		691
36.6 Posibilidad de atender al consumidor individualmente		692
36.7 Otro beneficio, cuál? (especifique):		693
36.8 No obtuvo ningún beneficio		694

37. Para las empresas que no hayan vendido por internet, ¿por qué razones no ha utilizado internet como herramienta de ventas?		
37.1 Falta de infraestructura/logística		803
37.2 Falta de confianza como herramienta de ventas		804
37.3 El cliente no utiliza esta herramienta		805
37.4 La empresa no tiene personal idóneo para utilizar esta herramienta		806
37.5 Otras razones (especifique):		807

38. ¿La empresa ordenó productos a través de Internet?		
Si		139
No		

39. ¿Qué tipo de acceso/ancho de banda utilizaba principalmente la empresa para acceder a Internet?		
39.1 Módem análogo/RDSI (ISDN) / dial up		141
39.2 XDSL (ADSL, SDSL, VDSL, etc.)		143
39.3 Cable / Fibra óptica		144
39.4 Inalámbrico móvil		145
39.5 Inalámbrico fijo (wi-fi)		146
39.6 Otro (especifique):		147

40. ¿Para qué servicios/actividades la empresa utilizaba Internet? (múltiples respuestas posibles)		
40.1 Comunicación (e-mail)		153
40.2 Búsqueda de información sobre productos y servicios		155
40.3 Búsqueda de información sobre organismos gubernamentales/autoridades públicas		156
40.4 Búsqueda de información sobre actividades de Investigación y Desarrollo		157
40.5 Banca electrónica y otros servicios financieros		159
40.6 Transacciones con organismos gubernamentales/autoridades públicas		160
40.7 Servicio al cliente		161
40.8 Distribución de productos en línea		162
40.9 Otras búsquedas de información		158

41. ¿La empresa contaba con una Red de Área Local (LAN)? (LAN: red que conecta computadoras que están dentro de un área localizada, como un edificio, un departamento, etc.; puede ser inalámbrica)		
Si		149
No		

x

42. ¿La empresa contaba con Intranet? (Intranet: se refiere a una red que utiliza el mismo protocolo que internet y que permite la comunicación dentro de una organización)		
Si		135
No		

x

43. ¿La empresa contaba con Extranet? (Extranet: trabaja con el protocolo de internet. Permite que usuarios externos seleccionados accedan a algunas partes de la red interna de una organización).		
Si		151
No		

x

44. ¿Disponía su empresa de alguna de las siguientes aplicaciones informáticas?		Si	No	930

¿Cuáles?



44.1 Gestionar información de clientes (herramientas CRM - Customer Relation Management) CRM: sistema de software de gestión basado en la atención al cliente que permite entablar una mayor proximidad e innovar la atención brindada al mismo.		695
44.2 Capturar, almacenar y compartir con otras áreas funcionales de la empresa información sobre clientes		696
44.3 Analizar la información disponible acerca de los clientes con fines comerciales y de marketing (fijación de precios, promociones comerciales, selección de canales de distribución, etc.)		697
44.4 Para compartir la información sobre compras y ventas con otras áreas funcionales de la empresa (herramientas ERP - Enterprise Resource Planning - para compartir información entre Finanzas, Organización, Marketing, etc.) ERP: sistemas de información gerenciales asociados a las operaciones de producción y distribución (ejemplo: logística, facturas, etc.)		698
44.5 Para dar soporte a las actividades de innovación (I+D, capacitación, ingeniería y diseño industrial, transferencia de tecnología, etc.)		699

by Carlo de Jorgue

SIGA

ACTIVIDADES DE MEDIO AMBIENTE EN EL AÑO 2016

45. Indique si la empresa ha realizado alguna/s de las siguientes acciones o prácticas.			
45.1	Incorporación de sistemas y/o equipos de tratamiento y/o disposición de efluentes líquidos y/o gaseosos y residuos sólidos		345
45.2	Remediación del medio ambiente		346
45.3	Mejoras en la eficiencia del uso del agua		347
45.4	Utilización eficiente de la energía		348
45.5	Optimización del uso de otros insumos		349
45.6	Reemplazó o modificó procesos o insumos contaminantes, por otros menos contaminantes.		350
45.7	Reconversión para lograr un "Producto" final menos contaminante		351
45.8	Reciclado o reutilización interno y/o externo de insumos y productos		352
45.9	Construcción de un mapa de riesgos ambientales de su actividad		353
45.10	Establecimiento de planes de prevención y mitigación de riesgos ambientales		354
45.11	Contrató algún seguro ambiental para las actividades de la empresa		355
45.12	Otras prácticas o acciones vinculadas a la gestión ambiental. Especifique		356
45.13	No ha realizado actividades relacionadas con el medio ambiente		357
Certificaciones relacionadas a la gestión ambiental			Certificaciones en proceso
45.14	ISO 14001	358.1	358.2
45.15	IRAM 3800	359.1	359.2
45.16	OHSAS	360.1	360.2
45.17	Otras. Especifique	361.1	361.2

X Debe elegir al menos una opción

Si en la pregunta 45 no ha desarrollado ninguna actividad, no deberá completar este cuadro. Pase al punto 47.

46. Indique cuáles han sido las motivaciones para desarrollar las acciones o prácticas mencionadas en la pregunta 45 (Considere que en la primer columna puede marcar más de una opción)		Motivaciones	Motivación principal (marcar sólo una opción de las seleccionadas en la columna anterior)
46.1	Reducir costos	379	390
46.2	Mejorar la imagen de la firma	380	391
46.3	Regulaciones (leyes, decretos, etc.) locales	381	392
46.4	Regulaciones (leyes, acuerdos, etc.) internacionales	382	393
46.5	Exigencias de clientes locales	383	394
46.6	Exigencias de mercados externos	384	395
46.7	Estándares intra-corporación	385	396
46.8	Para obtener certificaciones ambientales	386	397
46.9	Emular las acciones de competidores locales	387	398
46.10	Por exigencias de crédito (local o internacional)	388	399
46.11	Otros motivos. Especifique :	389	400

X

Para la realización de las actividades relacionadas con el Medio Ambiente, se debe recurrir a la utilización de distintas tecnologías modernas. En la pregunta 47 se solicita a la empresa, tanto si realizó como si no realizó actividades de protección ambiental, que identifique los obstáculos que la empresa sufriera para acceder a dichas tecnologías entre los distintos obstáculos propuestos.

Recuerde completar este cuadro aunque no haya realizado actividades en materia de protección ambiental

47. Indique cuáles son los obstáculos que ha debido enfrentar o enfrenta la empresa para acceder a nuevas tecnologías de protección del medio ambiente		Obstáculos	Obstáculo principal (marcar sólo una opción de las seleccionadas en la columna anterior)
47.1	Baja disponibilidad de dichas tecnologías en el mercado internacional	401	410
47.2	Baja disponibilidad de dichas tecnologías en el mercado local	402	411
47.3	Alto costo de las tecnologías disponibles	403	412
47.4	Las tecnologías existentes no son libremente accesibles por estar protegidas por patentes u otro tipo de mecanismos de propiedad intelectual	404	413
47.5	Carece de información sobre la existencia de nuevas tecnologías para la protección del medio ambiente	405	414
47.6	Falta de personal capacitado para su operación	406	415
47.7	Restricciones administrativas o regulatorias para la incorporación de tecnologías innovadoras o poco conocidas	407	416
47.8	Escasez de incentivos económicos para favorecer la adquisición de tecnología de última generación	408	417
47.9	Otros motivos. Especifique	409	418
47.10	No se presentaron obstáculos	084	

X

A2 Sintaxis procesamiento R

```
#####  
# AJUSTE DE MODELO LOGISTICO MIXTO  
# ESTIMACIONES DEL EFECTO ALEATORIO  
#####  
  
##### AJUSTE DEL MODELO (Base de Entrenamiento) #####  
  
attach(empresas) #empresas = base de entrenamiento  
  
library(lme4)  
  
modelo_log_mix<-glmer(INNPRODPROC_MDO ~ LN_RRHH +  
                    `I+DINTING_VTATOT`+ GTOINNEXT_VTATOT +  
                    IDINT_ING_CONT4+  
                    FONDOS_PROGOFIC + VINCULA +  
                    (1|NEMPRE), family=binomial, data=empresas,  
nAGQ=16,verbose=FALSE)  
summary(modelo_log_mix)  
  
##### Validacion del modelo - base de entrenamiento #####  
  
beta.est.m.l<-as.numeric(fixef(modelo_log_mix))  
aux.glmer<-as.numeric(VarCorr(modelo_log_mix)[[1]])  
sigma.est<-(aux.glmer[1])^0.5 # guarda ds ordenada aleatoria  
  
N<-nrow(empresas)  
total.grupos.trein<-nlevels(as.factor(NEMPRE))  
  
prob_est_modelo<-fitted(modelo_log_mix,se.fit=FALSE)  
  
library(PresenceAbsence)  
  
datos_m.l<-data.frame(NEMPRE,INNPRODPROC_MDO,prob_est_modelo) #banco  
de datos  
  
#Curva ROC modelo  
auc.roc.plot( datos_m.l,  
              threshold=101,  
              which.model=c(1),  
              model.names=c("modelo"),
```

```

na.rm=TRUE,
xlab="1-Especificidad (falsos positivos)",
ylab="Sensibilidad(verdaderos positivos)",
main="ROC Plot",
color=TRUE,
line.type=TRUE,
lwd=1,
mark=0,
mark.numbers=TRUE,
opt.thresholds=TRUE,
opt.methods = c(1,2,3,5),
req.sens=0.85,
req.spec=0.85,
obs.prev=NULL,
add.legend=TRUE,
legend.text=NULL,
legend.cex=0.7,
add.opt.legend=TRUE,
opt.legend.cex=0.7,
counter.diagonal=TRUE,
pch=NULL)

```

#Tabla de especificidad, sensibilidad, porcentaje de buena clasificacion

```

tabla_mconst<-
round(rbind(specificity(cmx(datos_m.1,which.model=1,threshold=0.27),s
t.dev=FALSE),

sensitivity(cmx(datos_m.1,which.model=1,threshold=0.27),st.dev=FALSE)
,

pcc(cmx(datos_m.1,which.model=1,threshold=0.27),st.dev=FALSE),
      auc(datos_m.1,
which.model=1,st.dev=FALSE)),4)
row.names(tabla_mconst)<-c("Especificidad", "Sensibilidad",
"Precisi n", "AUC")
colnames (tabla_mconst)<-c("Punto de corte Sens=Spec (0.27)")
tabla_mconst

```

Res duos quant licos aleatorizados

```

Y<-array(0,N)
for(j in 1:N) Y[j]<-rbinom(1,1,prob_est_modelo[j])

uij.a<-array(0,N)

```

```

uij.b<-array(0,N)
for(j in 1:N){
  uij.a[j]  <- pbinom(Y[j]-1,1, prob_est_modelo[j], lower.tail=TRUE)
  uij.b[j]  <- pbinom(Y[j],1, prob_est_modelo[j], lower.tail=TRUE)
}

u<-array(0,N)
for(j in 1:N) u[j] <- runif(n = length(Y[j]), min = uij.a[j], max =
uij.b[j])

rq<-array(0,N)
for(j in 1:N) rq[j]<-qnorm(u[j])

par(mfrow=c(1,3))
plot(rq, ylab="Quantile Residuals",ylim=c(-4,4), main="(a)")
qqnorm(rq,ylab="QQ-Plot: Quantile Residuals",ylim=c(-4,4),
main="(b)" )
qqline(rq)
qqnorm(ranef(modelo_log_mix)$NEMPRE[,1],ylab="QQ Plot: Random
Intercept",main="(c)")
qqline(ranef(modelo_log_mix)$NEMPRE[,1])

#test de normalidad de los residuos

shapiro.test(rq)  #H0= los datos son normales

##### ESTIMACION DE LOS EFECTOS ALEATORIOS #####

#####          Base de entrenamiento          #####

# Vector de caracteristicas agregadas por empresa + efectos aleatorios

vetor.caracteristica.trein<-matrix(0, total.grupos.trein,29)
for(k in 1: total.grupos.trein) {
  vetor.caracteristica.trein[k,29]<-ranef(modelo_log_mix)$NEMPRE[k,1]
}

#Creación de las variables agregadas a nivel de grupo (NEMPRE)

library(stats)

mean.trein.LN_RRHH      <-aggregate(      LN_RRHH      , by=list(NEMPRE),
      mean  )[,2]

```

```

mean.trein.LN_VTATOT <-aggregate( LN_VTATOT , by=list(NEMPRE),
  mean )[,2]
mean.trein.LN_PROD <-aggregate( LN_PROD , by=list(NEMPRE),
  mean )[,2]
mean.trein.PROF_RRHH <-aggregate( PROF_RRHH , by=list(NEMPRE),
  mean )[,2]
mean.trein.IDINT_ING_VTATOT <-aggregate( `I+DINTING_VTATOT` ,
  by=list(NEMPRE), mean )[,2]
mean.trein.GTOINNEXT_VTATOT <-aggregate( GTOINNEXT_VTATOT ,
  by=list(NEMPRE), mean )[,2]
mean.trein.EXPO_VTATOT <-aggregate( EXPO_VTATOT, by=list(NEMPRE),
  mean )[,2]

median.trein.LN_RRHH <-aggregate( LN_RRHH , by=list(NEMPRE),
  median )[,2]
median.trein.LN_VTATOT <-aggregate( LN_VTATOT , by=list(NEMPRE),
  median )[,2]
median.trein.LN_PROD <-aggregate( LN_PROD , by=list(NEMPRE),
  median )[,2]
median.trein.PROF_RRHH <-aggregate( PROF_RRHH , by=list(NEMPRE),
  median )[,2]
median.trein.IDINT_ING_VTATOT <-aggregate( `I+DINTING_VTATOT` ,
  by=list(NEMPRE), median )[,2]
median.trein.GTOINNEXT_VTATOT <-aggregate( GTOINNEXT_VTATOT ,
  by=list(NEMPRE), median )[,2]
median.trein.EXPO_VTATOT <-aggregate( EXPO_VTATOT,
  by=list(NEMPRE), median )[,2]

min.trein.LN_RRHH <-aggregate( LN_RRHH , by=list(NEMPRE),
  min )[,2]
min.trein.LN_VTATOT <-aggregate( LN_VTATOT , by=list(NEMPRE),
  min )[,2]
min.trein.LN_PROD <-aggregate( LN_PROD , by=list(NEMPRE),
  min )[,2]
min.trein.PROF_RRHH <-aggregate( PROF_RRHH , by=list(NEMPRE),
  min )[,2]
min.trein.IDINT_ING_VTATOT <-aggregate( `I+DINTING_VTATOT` ,
  by=list(NEMPRE), min )[,2]
min.trein.GTOINNEXT_VTATOT <-aggregate( GTOINNEXT_VTATOT ,
  by=list(NEMPRE), min )[,2]
min.trein.EXPO_VTATOT <-aggregate( EXPO_VTATOT, by=list(NEMPRE),
  min )[,2]

max.trein.LN_RRHH <-aggregate( LN_RRHH , by=list(NEMPRE),
  max )[,2]
max.trein.LN_VTATOT <-aggregate( LN_VTATOT , by=list(NEMPRE),
  max )[,2]
max.trein.LN_PROD <-aggregate( LN_PROD , by=list(NEMPRE),
  max )[,2]

```

```

max.trein.PROF_RRHH <-aggregate( PROF_RRHH , by=list(NEMPRE),
max ) [,2]
max.trein.IDINT_ING_VTATOT <-aggregate( `I+DINTING_VTATOT` ,
by=list(NEMPRE), max ) [,2]
max.trein.GTOINNEXT_VTATOT <-aggregate( GTOINNEXT_VTATOT ,
by=list(NEMPRE), max ) [,2]
max.trein.EXPO_VTATOT <-aggregate( EXPO_VTATOT , by=list(NEMPRE),
max ) [,2]

vetor.caracteristica.trein<-matrix(0, total.grupos.trein,29)
for(k in 1: total.grupos.trein){
vetor.caracteristica.trein[k,1]<-mean.trein.LN_RRHH[k]
vetor.caracteristica.trein[k,2]<-mean.trein.LN_VTATOT[k]
vetor.caracteristica.trein[k,3]<-mean.trein.LN_PROD[k]
vetor.caracteristica.trein[k,4]<-mean.trein.PROF_RRHH[k]
vetor.caracteristica.trein[k,5]<-mean.trein.IDINT_ING_VTATOT[k]
vetor.caracteristica.trein[k,6]<-mean.trein.GTOINNEXT_VTATOT[k]
vetor.caracteristica.trein[k,7]<-mean.trein.EXPO_VTATOT[k]

vetor.caracteristica.trein[k,8]<-median.trein.LN_RRHH[k]
vetor.caracteristica.trein[k,9]<-median.trein.LN_VTATOT[k]
vetor.caracteristica.trein[k,10]<-median.trein.LN_PROD[k]
vetor.caracteristica.trein[k,11]<-median.trein.PROF_RRHH[k]
vetor.caracteristica.trein[k,12]<-median.trein.IDINT_ING_VTATOT[k]
vetor.caracteristica.trein[k,13]<-median.trein.GTOINNEXT_VTATOT[k]
vetor.caracteristica.trein[k,14]<-median.trein.EXPO_VTATOT[k]

vetor.caracteristica.trein[k,15]<-min.trein.LN_RRHH[k]
vetor.caracteristica.trein[k,16]<-min.trein.LN_VTATOT[k]
vetor.caracteristica.trein[k,17]<-min.trein.LN_PROD[k]
vetor.caracteristica.trein[k,18]<-min.trein.PROF_RRHH[k]
vetor.caracteristica.trein[k,19]<-min.trein.IDINT_ING_VTATOT[k]
vetor.caracteristica.trein[k,20]<-min.trein.GTOINNEXT_VTATOT[k]
vetor.caracteristica.trein[k,21]<-min.trein.EXPO_VTATOT[k]

vetor.caracteristica.trein[k,22]<-max.trein.LN_RRHH[k]
vetor.caracteristica.trein[k,23]<-max.trein.LN_VTATOT[k]
vetor.caracteristica.trein[k,24]<-max.trein.LN_PROD[k]
vetor.caracteristica.trein[k,25]<-max.trein.PROF_RRHH[k]
vetor.caracteristica.trein[k,26]<-max.trein.IDINT_ING_VTATOT[k]
vetor.caracteristica.trein[k,27]<-max.trein.GTOINNEXT_VTATOT[k]
vetor.caracteristica.trein[k,28]<-max.trein.EXPO_VTATOT[k]

vetor.caracteristica.trein[k,29]<-ranef(modelo_log_mix)$NEMPRE[k,1]

```

```

}
detach(package:lme4)

#####          Base de validacion          #####

detach()

attach(empresas.teste)      #empresas.teste = base de validacion

total.grupo.teste<- nrow(empresas.teste)
n.teste<-nrow(empresas.teste)

# Creacion de variables agregadas a nivel de grupo

library(stats)

mean.teste.LN_RRHH      <-aggregate(      LN_RRHH      , by=list(NEMPRE),
      mean      )[,2]
mean.teste.LN_VTATOT    <-aggregate(      LN_VTATOT    , by=list(NEMPRE),
      mean      )[,2]
mean.teste.LN_PROD      <-aggregate(      LN_PROD      , by=list(NEMPRE),
      mean      )[,2]
mean.teste.PROF_RRHH    <-aggregate(PROF_RRHH , by=list(NEMPRE),
      mean      )[,2]
mean.teste.IDINT_ING_VTATOT <-aggregate(`I+DINTING_VTATOT`      ,
      by=list(NEMPRE), mean      )[,2]
mean.teste.GTOINNEXT_VTATOT <-aggregate(GTOINNEXT_VTATOT ,
      by=list(NEMPRE), mean      )[,2]
mean.teste.EXPO_VTATOT <-aggregate( EXPO_VTATOT      , by=list(NEMPRE),
      mean      )[,2]

median.teste.LN_RRHH    <-aggregate(      LN_RRHH      , by=list(NEMPRE),
      median      )[,2]
median.teste.LN_VTATOT <-aggregate(      LN_VTATOT    , by=list(NEMPRE),
      median      )[,2]
median.teste.LN_PROD    <-aggregate(      LN_PROD      , by=list(NEMPRE),
      median      )[,2]
median.teste.PROF_RRHH <-aggregate(      PROF_RRHH    , by=list(NEMPRE),
      median      )[,2]
median.teste.IDINT_ING_VTATOT <-aggregate(      `I+DINTING_VTATOT`      ,
      by=list(NEMPRE), median      )[,2]
median.teste.GTOINNEXT_VTATOT <-aggregate(      GTOINNEXT_VTATOT ,
      by=list(NEMPRE), median      )[,2]
median.teste.EXPO_VTATOT <-aggregate( EXPO_VTATOT      ,
      by=list(NEMPRE), median      )[,2]

```

```

min.teste.LN_RRHH      <-aggregate(      LN_RRHH      , by=list(NEMPRE),
  min      )[,2]
min.teste.LN_VTATOT   <-aggregate(      LN_VTATOT   , by=list(NEMPRE),
  min      )[,2]
min.teste.LN_PROD     <-aggregate(      LN_PROD     , by=list(NEMPRE),
  min      )[,2]
min.teste.PROF_RRHH   <-aggregate(      PROF_RRHH   , by=list(NEMPRE),
  min      )[,2]
min.teste.IDINT_ING_VTATOT <-aggregate(      `I+DINTING_VTATOT`,
  by=list(NEMPRE), min      )[,2]
min.teste.GTOINNEXT_VTATOT <-aggregate(      GTOINNEXT_VTATOT ,
  by=list(NEMPRE), min      )[,2]
min.teste.EXPO_VTATOT <-aggregate(      EXPO_VTATOT , by=list(NEMPRE),
  min      )[,2]

max.teste.LN_RRHH     <-aggregate(      LN_RRHH     , by=list(NEMPRE),
  max      )[,2]
max.teste.LN_VTATOT   <-aggregate(      LN_VTATOT   , by=list(NEMPRE),
  max      )[,2]
max.teste.LN_PROD     <-aggregate(      LN_PROD     , by=list(NEMPRE),
  max      )[,2]
max.teste.PROF_RRHH   <-aggregate(      PROF_RRHH   , by=list(NEMPRE),
  max      )[,2]
max.teste.IDINT_ING_VTATOT <-aggregate(      `I+DINTING_VTATOT`      ,
  by=list(NEMPRE), max      )[,2]
max.teste.GTOINNEXT_VTATOT <-aggregate(      GTOINNEXT_VTATOT ,
  by=list(NEMPRE), max      )[,2]
max.teste.EXPO_VTATOT <-aggregate(      EXPO_VTATOT , by=list(NEMPRE),
  max      )[,2]

vetor.caracteristica.teste<-array(0,c(total.grupo.teste,29))
for(k in 1: total.grupo.teste){
  vetor.caracteristica.teste[k,1]<-mean.teste.LN_RRHH[k]
  vetor.caracteristica.teste[k,2]<-mean.teste.LN_VTATOT[k]
  vetor.caracteristica.teste[k,3]<-mean.teste.LN_PROD[k]
  vetor.caracteristica.teste[k,4]<-mean.teste.PROF_RRHH[k]
  vetor.caracteristica.teste[k,5]<-mean.teste.IDINT_ING_VTATOT[k]
  vetor.caracteristica.teste[k,6]<-mean.teste.GTOINNEXT_VTATOT[k]
  vetor.caracteristica.teste[k,7]<-mean.teste.EXPO_VTATOT[k]

  vetor.caracteristica.teste[k,8]<-median.teste.LN_RRHH[k]
  vetor.caracteristica.teste[k,9]<-median.teste.LN_VTATOT[k]
  vetor.caracteristica.teste[k,10]<-median.teste.LN_PROD[k]
  vetor.caracteristica.teste[k,11]<-median.teste.PROF_RRHH[k]
  vetor.caracteristica.teste[k,12]<-median.teste.IDINT_ING_VTATOT[k]
}

```

```

vetor.caracteristica.teste[k,13]<-median.teste.GTOINNEXT_VTATOT[k]
vetor.caracteristica.teste[k,14]<-median.teste.EXPO_VTATOT[k]

vetor.caracteristica.teste[k,15]<-min.teste.LN_RRHH[k]
vetor.caracteristica.teste[k,16]<-min.teste.LN_VTATOT[k]
vetor.caracteristica.teste[k,17]<-min.teste.LN_PROD[k]
vetor.caracteristica.teste[k,18]<-min.teste.PROF_RRHH[k]
vetor.caracteristica.teste[k,19]<-min.teste.IDINT_ING_VTATOT[k]
vetor.caracteristica.teste[k,20]<-min.teste.GTOINNEXT_VTATOT[k]
vetor.caracteristica.teste[k,21]<-min.teste.EXPO_VTATOT[k]

vetor.caracteristica.teste[k,22]<-max.teste.LN_RRHH[k]
vetor.caracteristica.teste[k,23]<-max.teste.LN_VTATOT[k]
vetor.caracteristica.teste[k,24]<-max.teste.LN_PROD[k]
vetor.caracteristica.teste[k,25]<-max.teste.PROF_RRHH[k]
vetor.caracteristica.teste[k,26]<-max.teste.IDINT_ING_VTATOT[k]
vetor.caracteristica.teste[k,27]<-max.teste.GTOINNEXT_VTATOT[k]
vetor.caracteristica.teste[k,28]<-max.teste.EXPO_VTATOT[k]

}

#### METODOS DE PREDICCION PARA NUEVOS GRUPOS #####

##### METODO EBP #####

tam.grupo.novo<-0
count<-1
tam.grupo.novo[1]<-1
for(i in 2:n.teste) {
  if(NEMPRE[i] != NEMPRE[i-1]){
    count <- count + 1
    tam.grupo.novo[count] <- 0
  }
  tam.grupo.novo[count] <- 1 + tam.grupo.novo[count]
}
total.grupos.val.novo<-count

yi.novo<-array(0,total.grupos.val.novo)
for (i in 1: total.grupos.val.novo){
  yi.novo[i]<-tam.grupo.novo[i]/2
}

ef.fixo.laplace<- array(0, n.teste)

```

```

for(i in 1 : total.grupo.teste){
  for(j in 1: n.teste){
    ef.fixo.laplace[j]<-
(1*beta.est.m.l[1]+beta.est.m.l[2]*empresas.teste$LN_RRHH[j]+
beta.est.m.l[3]*empresas.teste$I+DINTING_VTATOT`[j]+
beta.est.m.l[4]*empresas.teste$GTOINNEXT_VTATOT[j]+
beta.est.m.l[5]*empresas.teste$IDINT_ING_CONT4[j]+
beta.est.m.l[6]*empresas.teste$FONDOS_PROGOFIC[j]+
beta.est.m.l[7]*empresas.teste$VINCULA[j])
  }
}

sigma<-1
mu<-0

#variables auxiliares para prediccion del modelo mixto
partel<-array(0,c(n.teste,1))
parte2<-array(0,c(n.teste,1))

laplace.teste.ebp<- array(0, n.teste)

for(i in 1: (total.grupo.teste)){
  ii <- (i-1)*tam.grupo.novo[i]

  texto_somatoria <- "("
  for (qq in 1:tam.grupo.novo[i])
    texto_somatoria <-
paste(texto_somatoria,"+log(1+exp(ef.fixo.laplace[ii+",as.character(q
q),")+sigma.est*z))")
  texto_somatoria <- paste(texto_somatoria,")")

  for(k in 1:tam.grupo.novo[i]) {
    j <- ii+k
    texto_integrand <- paste("function(z)
      (
      (
      exp(
      (yi.novo[i]+1) * sigma.est * z
      -(",
      texto_somatoria, ")

```

```

)
/
(
  1+exp(
    ef.fixo.laplace[j]+sigma.est*z
  )
)
)
)
*
(
  ( 1/( sqrt(2*pi)*sigma ) )
)
*
exp(
  -( (z - mu)^2 / (2*sigma^2) )
)
)
) " )

integrand <- eval(parse(text=texto_integrand))

aux1<-integrate(integrand, lower = -Inf, upper =Inf)

partel[j]<-aux1$value

texto_integrand2 <- paste("function(z)
  (
    (
      exp(
        yi.novo[i]*sigma.est*z
        -(",
      texto_somatoria, ")
    )
  )
  *
  (
    ( 1/(sqrt(2*pi)*sigma) )
  )
  *
  exp(
    -( (z - mu)^2/(2*sigma^2) )
  )
)
) " )

integrand2 <- eval(parse(text=texto_integrand2))

```

```

        aux2<-integrate(integrand2, lower = -Inf, upper =Inf)
    parte2[j]<-aux2$value
    laplace.teste.ebp[j]<-exp(ef.fixo.laplace[j])*(partel[j]/parte2[j])
}
}

##### METODO NAIVE #####

laplace.teste.naive<- array(0, n.teste)

for(i in 1 : total.grupo.teste){
    for(j in 1: n.teste){

        # parte aleatoria anulada
        laplace.teste.naive[j]<-
exp(ef.fixo.laplace[j])/(1+exp(ef.fixo.laplace[j]))
    }
}

##### METODO REGRESION LINEAL #####

library(MASS)

#prueba para elegir el modelo de regresion
#columnas de medias: c(1,2,3,4,5,6,7), medianas:
c(8,9,10,11,12,13,14) , min: c(15,16,17,18,19,20,21), max:
c(22,23,24,25,26,27,28)
mod_medias <- lm(vetor.caracteristica.trein[,29] ~ ., data =
as.data.frame(vetor.caracteristica.trein[,c(1,2,3,4,5,6,7)])) #dejar
solo las columnas de la base donde estan las vbles para probar en el
modelo
stepAIC(mod_medias)

mod_medianas <- lm(vetor.caracteristica.trein[,29] ~ ., data =
as.data.frame(vetor.caracteristica.trein[,c(8,9,10,11,12,13,14)]))
#dejar solo las columnas de la base donde estan las vbles para probar
en el modelo
stepAIC(mod_medianas)

mod_min <- lm(vetor.caracteristica.trein[,29] ~ ., data =
as.data.frame(vetor.caracteristica.trein[,c(15,16,17,18,19,20,21)]))
#dejar solo las columnas de la base donde estan las vbles para probar
en el modelo
stepAIC(mod_min)

```

```

mod_max <- lm(vetor.caracteristica.trein[,29] ~ ., data =
as.data.frame(vetor.caracteristica.trein[,c(22,23,24,25,26,27,28)]))
#dejar solo las columnas de la base donde estan las vbles para probar
en el modelo
stepAIC(mod_max)

#modelo elegido
#medidas agrupas segun la mediana, explicativas GTOINNEXT_VTA -
#modelo con menor AIC
fit.intecepto.al.laplace<-lm(vetor.caracteristica.trein[,29]~
1+vetor.caracteristica.trein[,c(13)])

summary(fit.intecepto.al.laplace)

beta.lr.0<-as.numeric(fit.intecepto.al.laplace$coefficients[1])
beta.lr.1<-as.numeric(fit.intecepto.al.laplace$coefficients[2])

ef.al.lr<- array(0, n.teste)
for(i in 1 : total.grupo.teste){
  for(j in 1: n.teste){
    ef.al.lr[j]<-
(1*beta.lr.0+beta.lr.1*empresas.teste$GTOINNEXT_VTATOT[j])

  }
}

laplace.teste.lr<- array(0, n.teste)

for(i in 1 : total.grupo.teste){
  for(j in 1: n.teste){

    laplace.teste.lr[j]<-
exp(ef.fixo.laplace[j]+ef.al.lr[j]) / (1+exp(ef.fixo.laplace[j]+ef.al.l
r[j]))

  }
}

##### METODO DE VECINO MAS CERCANO #####

lambda<- 7
qtde_caracteristica<-7 #medidas resumen y variables utilizadas para
los calculos

# entrenamiento

```

```

vp <- vetor.caracteristica.trein[,c(8,9,10,11,12,13,14)]

vi <- vetor.caracteristica.teste[,1:qtde_caracteristica] #validación

np = dim(vp) [1]
ni = dim(vi) [1]
nc = dim(vp) [2]
INF <- 999999
NIL <- -1

# Matriz con los indices de empresa en filas e indices de distancias
ordenadas en columnas
nn = matrix(NIL,ni,np)
nn_dist = matrix(NIL,ni,np)

# vector de distancias
vd <- c(rep(NIL,np))
S <- cov(vp)
for(ii in 1:ni) {
  for(jj in 1:np) {
    # distancia
    dist <- 0
    dist.num<-0
    dist.den<-0

    #Distancia Minkowski
    tmp_dif <- abs(mean(vi[ii,]) - mean(vp[jj,]))
    dist <- dist + tmp_dif^lambda
    dist<-dist^(1/lambda)

    vd[jj] <- dist
  }
  nn[ii,] <- sort(vd, index.return = TRUE)$ix
  nn_dist[ii,] <- sort(vd, index.return = FALSE)
}

auc.values<-array(0,np)
ks.values<-array(0,np)

for(k.nn in 1:np){
  totknn <- k.nn

```

```

for(ii in 1:ni){

# Ponderacion exponencial de las distancias
  tmp1 <- 0
  tmp2 <- 0
  tmp_dn <- 0
  for(kk in 1:totknn) {
    jj <- nn[ii,kk]
    tmp_d <- nn_dist[ii,kk]
    tmp_dn <- tmp_dn + exp(-tmp_d)
    tmp1<-tmp1+vector.caracteristica.trein[jj,29]*exp(-tmp_d)

  }
  tmp_efeito1<-tmp1/tmp_dn

# Medoid
  tmp_e1 <- tmp_efeito1
  tmp_min_dist <- INF
  tmp_min_ind <- NIL
  for(kk in 1:totknn) {
    jj <- nn[ii,kk]
    tmp_dif1 <- tmp_e1 - vector.caracteristica.trein[jj,29]
    dist <- tmp_dif1*tmp_dif1
    # elemento con menor distancia hasta el momento
    if (dist < tmp_min_dist) {
      tmp_min_dist <- dist
      tmp_min_ind <- jj
    }
  }
  tmp_efeito1<-vector.caracteristica.trein[tmp_min_ind,29]

  vector.caracteristica.teste[ii,29]<-tmp_efeito1

}
}

laplace.teste.knn<- array(0, n.teste)

for(i in 1 : total.grupo.teste){
  for(j in 1: n.teste){

```

```

    laplace.teste.knn[j]<-
exp(ef.fixo.laplace[j]+vetor.caracteristica.teste[i,29])/(1+exp(ef.fixo.laplace[j]+vetor.caracteristica.teste[i,29]))
  }
}

##### Comparacion de metodos de prediccion #####
#### Tabla de especificidad, sensibilidad, porcentaje de buena
#clasificacion

library(PresenceAbsence)

#banco de datos
datos_valid<- data.frame(empresas.teste$NEMPRE,
empresas.teste$INNPRODPROC_MDO,
                        laplace.teste.naive,
                        laplace.teste.ebp,
                        laplace.teste.lr,
                        laplace.teste.knn)

#punto de corte: ver roc del modelo de construcción

t_naive<-
round(rbind(specificity(cmx(datos_valid,which.model=1,threshold=0.27)
,st.dev=FALSE),

sensitivity(cmx(datos_valid,which.model=1,threshold=0.27),st.dev=FALSE),

pcc(cmx(datos_valid,which.model=1,threshold=0.27),st.dev=FALSE)),4)

t_ebp<-
round(rbind(specificity(cmx(datos_valid,which.model=2,threshold=0.27)
,st.dev=FALSE),

sensitivity(cmx(datos_valid,which.model=2,threshold=0.27),st.dev=FALSE),

pcc(cmx(datos_valid,which.model=2,threshold=0.27),st.dev=FALSE)),4)

t_lr<-
round(rbind(specificity(cmx(datos_valid,which.model=3,threshold=0.27)
,st.dev=FALSE),

sensitivity(cmx(datos_valid,which.model=3,threshold=0.27),st.dev=FALSE),

pcc(cmx(datos_valid,which.model=3,threshold=0.27),st.dev=FALSE)),4)

```

```

t_knn<-
round(rbind(specificity(cmx(datos_valid,which.model=4,threshold=0.27)
,st.dev=FALSE),

sensitivity(cmx(datos_valid,which.model=4,threshold=0.27),st.dev=FALSE),

pcc(cmx(datos_valid,which.model=4,threshold=0.27),st.dev=FALSE)),4)

tabla_mvalid<-cbind(t_naive,t_ebp, t_lr,t_knn)
row.names(tabla_mvalid)<-c("Especificidad", "Sensibilidad",
"Precisi n")
colnames(tabla_mvalid)<-c("Naive", "EBP", "LR", "KNN")
tabla_mvalid

## Se guardan probabilidades estimadas segun cada metodo #####

base.teste.probab<-data.frame(empresas.teste$NEMPRE,
empresas.teste$ANO,
empresas.teste$LN_RRHH,
empresas.teste$GTOINN_VTATOT,
empresas.teste$EXPO_VTATOT,
empresas.teste$FONDOS_PROGOFIC,
empresas.teste$VINC_PUB,
empresas.teste$INNPORDPROC_MDO,
laplace.teste.naive,
laplace.teste.knn,
laplace.teste.ebp,
laplace.teste.lr)

write.table(base.teste.probab, "D:/Tesis/Base y modelos
tesis/base.teste.probab.txt", sep="\t")
#####

```

NOTA: Sintaxis adaptada en base a los trabajos de Tamura, K.A. y Giampaoli, V. (2011, 2013) y Tamura, K.A., Giampaoli, V. y Noma, A. (2013).