

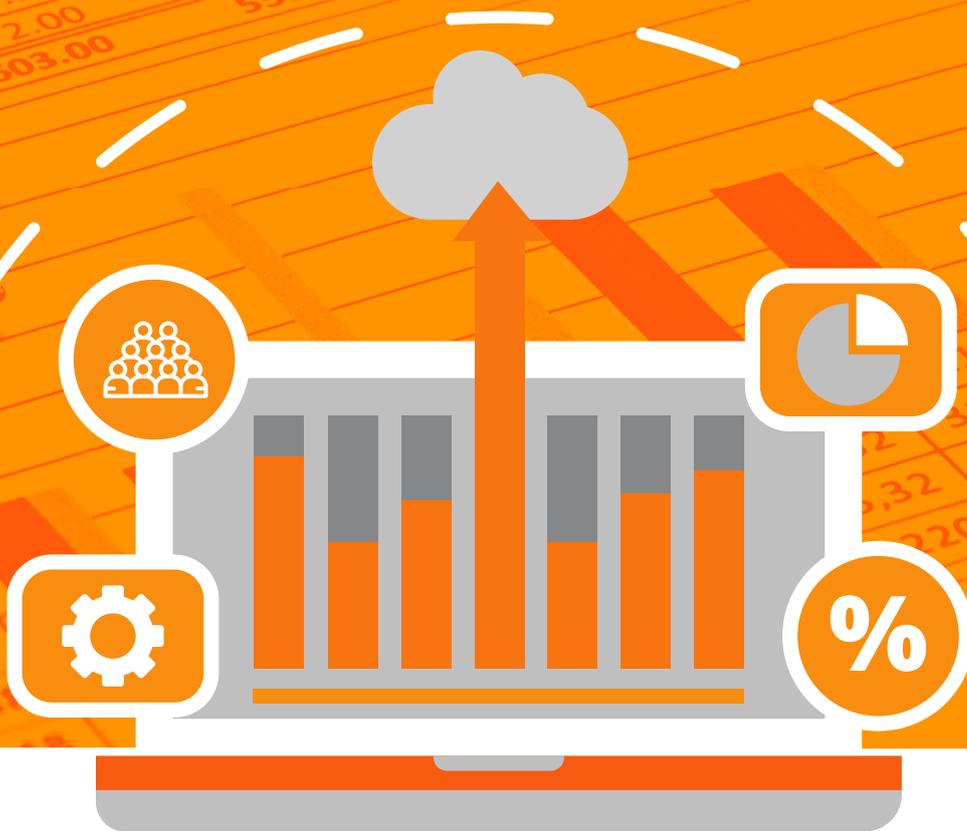
# DISSERTATIO

*Estadística*

COLEGIO DE GRADUADOS DE CIENCIAS ECONÓMICAS DE ROSARIO  
CONSEJO PROFESIONAL DE CIENCIAS ECONÓMICAS  
DE LA PROVINCIA DE SANTA FE CÁMARA II  
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

## TRABAJOS FINALES

RECENSIÓN DE TESIS Y PRÁCTICAS PROFESIONALES  
DE LA CARRERA LICENCIATURA EN ESTADÍSTICA



CONSEJO PROFESIONAL  
DE CIENCIAS ECONÓMICAS  
DE LA PROVINCIA DE SANTA FE



Colegio de Graduados  
en Ciencias Económicas  
de Rosario

# ÍNDICE

**CONFORMACIONES** 01

---

**UNA APUESTA CONJUNTA AL CRECIMIENTO PROFESIONAL** 02

---

**COMITÉ EDITORIAL** 03

---

## ARTÍCULOS

**MODELO DE CLASIFICACIÓN DE AFILIADOS  
PARA LA RETENCIÓN DE BAJAS** 04  
LIC. FLORENTINA BURLON

---

**ESTIMACIÓN EN DISEÑOS MUESTRALES COMPLEJOS  
EN PRESENCIA DE NO RESPUESTA** 13  
LIC. TOMÁS CAPRETTO

---

**ESTRATEGIAS PARA LA CONSTRUCCIÓN DE DISEÑOS  $2^{k-p}$  PLEGADOS  
DE ACUERDO A LOS CRITERIOS DE RESOLUCIÓN Y ABERRACIÓN** 23  
LIC. ANTONELLA N. MAZZEO

---

**ANÁLISIS DEL TIEMPO HASTA LA CURACIÓN DE ÚLCERAS  
VENOSAS MEDIANTE MODELOS LINEALES GENERALIZADOS  
PARA DATOS CENSURADOS** 33  
LIC. JULIA MILLET

---

**CONSTRUCCIÓN Y ANÁLISIS DE UN DISEÑO  $\alpha$  PARA  
LA COMPARACIÓN DE CULTIVARES DE TRIGO** 42  
LIC. CAMILA MIOTTI

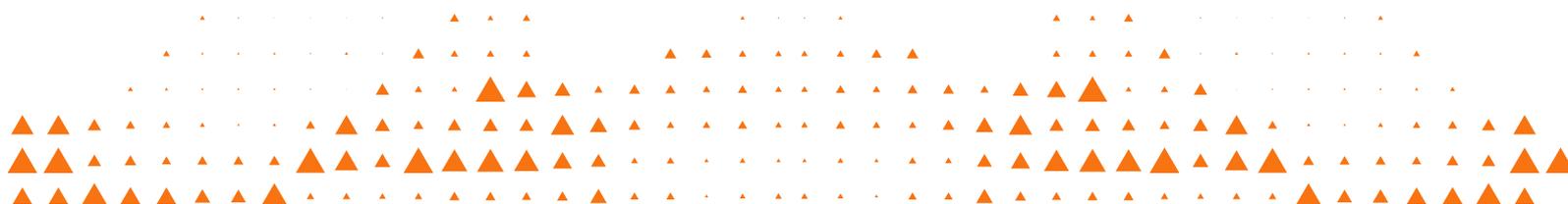
---

**MÉTODOS DE ESTIMACIÓN DE LA HETEROGENEIDAD ENTRE ESTUDIOS  
EN META-ANÁLISIS DE VARIABLES CONTINUAS** 52  
LIC. SOFÍA TAMUCH

---

**VECTORES DE CAPACIDAD DE PROCESOS MULTIVARIADOS.  
UN ESTUDIO COMPARATIVO** 63  
LIC. LUISINA VAZQUEZ

---



# CONFORMACIONES

## COMITÉ DIRECTIVO

Lic. Adriana Racca (FCEyE)  
Dr. Sergio M. Roldán (CPCE)  
Dra. Lidia Giovannoni (CGCE)

## COMITÉ ACADÉMICO

Mg. María Teresa Blaconá (FCEyE)  
Mg. Cristina Beatriz Cuesta (FCEyE)  
Dra. Marta Beatriz Quaglino (FCEyE)  
Lic. Nora Ventroni (Comisión de Estadística CPCE-CGCE)

## COMITÉ ACADÉMICO

Lic. Laura Balparda (Comisión de Estadística del CPCE-CGCE)  
Lic. Dalila Vadell (Comisión de Estadística del CGCE-CPCE)  
Dra. Lucía Hernandez (FCEyE)  
Lic. Cecilia Rapelli (FCEyE)

Esta revista se pone a disposición de los profesionales matriculados al Consejo Profesional de Ciencias Económicas de la Provincia de Santa Fe Cámara II (CPCE), asociados del Colegio de Graduados en Ciencias Económicas de Rosario (CGCE), estudiantes y docentes de la Facultad de Ciencias Económicas y Estadística (FCEyE) de la Universidad Nacional de Rosario (UNR) y otras Instituciones vinculadas al quehacer profesional y académico.

Su contenido puede ser reproducido en forma parcial o total citando la fuente. En caso de utilización deberá enviar dos ejemplares de la publicación respectiva a **Maipú 1344 – 2000 Rosario Tel. 4772727 email: [consejo@cpcesfe2.org.ar](mailto:consejo@cpcesfe2.org.ar)**

El contenido de los trabajos finales no necesariamente refleja la opinión de los Comités responsables de esta publicación digital.

Las Instituciones no son responsables por el contenido de las informaciones y opiniones que viertan en esta revista quienes son identificados como autores de dichos trabajos finales, en todos los casos deberán ser cotejadas por los Profesionales y/o las fuentes.





## UNA APUESTA CONJUNTA AL CRECIMIENTO PROFESIONAL

En un año tan particular signado por una pandemia mundial y sus respectivas consecuencias locales, todas las actividades, laborales, académicas, sociales, familiares, se resignifican y adquieren un nuevo matiz.

En este contexto, el trabajo conjunto entre instituciones requiere un mayor esfuerzo y pone de manifiesto la relevancia de su valor agregado y su compromiso permanente en pos de los profesionales.

Es por ello que se decidió continuar con las revistas digitales, elaboradas conjuntamente entre el Consejo Profesional en Ciencias Económicas de la Provincia de Santa Fe Cámara II, el Colegio de Graduados en Ciencias Económicas de Rosario y la Facultad de Cien-

cias Económicas y Estadística de la Universidad Nacional de Rosario.

Este proyecto comenzó en el año 2015 con *Dissertatio Economía*, en 2017 se sumó *Dissertatio Estadística* y en 2018 *Dissertatio Administración*, consolidando la participación de las tres escuelas de las licenciaturas que integran la facultad.

Estas revistas compendian una selección de tesinas de grado y trabajos finales, procurando incentivar el trabajo de investigación y facilitando el tránsito de la vida académica a la vida profesional, en la que la actualización y capacitación deberá estar presente siempre, siendo herramientas fundamentales para el desarrollo profesional.



En esta cuarta edición de *Dissertatio Estadística* (UNR-CGCE-CPCE) se presenta la recensión de seis tesinas y de un informe final de práctica profesional, realizadas por estudiantes de la Licenciatura en Estadística para cumplimentar el último requisito obligatorio para alcanzar su titulación. Las mismas se elaboran bajo la dirección y asesoramiento de docentes de la Escuela de Estadística de la Facultad de Ciencias Económicas y Estadística (FCEyE) de la Universidad Nacional de Rosario (UNR).

La obtención del título a través de una práctica profesional desarrollada en el ámbito laboral, es una modalidad que se implementó en el año 2017 y donde el futuro egresado tiene la oportunidad de contar, no sólo con la tutoría de un docente de la Escuela de Estadística (FCEyE-UNR), sino también con la inestimable guía de un Tutor Estadístico en ejercicio de su actividad profesional.

Al igual que en las ediciones anteriores, los temas seleccionados y las metodologías aplicadas constituyen nuevos desafíos para los autores, así como también para los docentes y tutores que tuvieron la oportunidad de guiar y supervisar estos trabajos, los cuales evidencian muchas horas de esfuerzo y dedicación.

Una vez más, con esta publicación digital, las instituciones participantes, a saber, Escuela de Estadística de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario, el Colegio de Graduados en Ciencias Económicas de Rosario y el Consejo Profesional en Ciencias Económicas de la Provincia de Santa Fe -Cámara II-, demuestran el fuerte interés por sostener y fortalecer un espacio de difusión técnico estadístico, de acceso libre y gratuito al público en general.

Cabe señalar que la edición completa de las tesinas se encuentra disponible en <https://www.fcecon.unr.edu.ar/web-nueva/tesinas-aprobadas> y los informes finales de las prácticas profesionales en <https://www.fcecon.unr.edu.ar/web-nueva/practicas-profesionales-aprobadas>.

Desde el Comité Editorial aprovechamos la oportunidad para agradecer y felicitar a todas las personas que hicieron posible esta producción y expresar nuestro fuerte deseo que el próximo año nos encuentre institucionalmente más comprometidos y juntos acompañando una nueva edición de la *Dissertatio Estadística*.

# MODELO DE CLASIFICACIÓN DE AFILIADOS PARA LA RETENCIÓN DE BAJAS

LIC. FLORENTINA BURLON

Tutora en la Escuela de Estadística: MG. CRISTINA CUESTA

Tutora en la Organización: LIC. ANABEL CHIAPPELLO

---

En este trabajo se estudia la retención de afiliados de una empresa de medicina prepaga mediante la aplicación de soluciones innovadoras que permiten generar un modelo de clasificación para predecir la posibilidad de que un afiliado decida darse de baja de la cobertura.

El alcance de esta propuesta abarca el comportamiento de todos los socios que estuvieron activos en algún momento entre el 01/01/2017 y 31/12/2018.

Se utilizan distintos métodos de aprendizaje supervisado: regresión logística, árboles de clasificación, *bagging* y *random forests*. Se ajustan modelos que relacionan la respuesta con variables predictoras, con el objetivo de predecir correctamente la respuesta para futuras observaciones y tener un mejor entendimiento de la relación entre estas variables.

Las características más asociadas a la “Baja” resultaron ser el tipo de Credencial (Integral, Superior, Selecta y Otras) y la antigüedad en la empresa desde la fecha de contratación. Sin embargo, los resultados obtenidos no son suficientemente contundentes para dar respuesta al objetivo del estudio. Si bien se detectaron altos porcentajes de clasificación correcta de la variable respuesta con la mayoría de las técnicas, ninguna arrojó un resultado de sensibilidad aceptable. El porcentaje más alto de buena predicción se obtuvo con la técnica de *bagging*.



## INTRODUCCIÓN

El presente informe es el resultado de la realización de una Práctica Profesional de la carrera Licenciatura en Estadística que fue llevada a cabo entre los meses de abril y junio de 2019 en la organización médico asistencial Aca Salud.

Si se analiza la demanda de servicio del sistema sanitario argentino es posible distinguir dos grandes grupos. Por un lado, aquellos que sólo acceden a los servicios públicos y, por otro, quienes además tienen cobertura privada. En Argentina, es inexacto afirmar que hay población con y sin cobertura de salud, porque el subsistema público tiene un alcance universal. No obstante, es consistente afirmar que hay población con y sin cobertura prepaga de salud. Según el Instituto Nacional de Estadística y Censos (INDEC), el 36% de los habitantes accede solamente a servicios públicos, un 16% cuenta con una prepaga como lo es Aca Salud y el resto accede al sistema de seguridad social (INDEC, 2010).

El mercado de las prepagas en Argentina está limitado, crecer en cantidad de socios es una tarea ardua que puede lograrse, por ejemplo, atrayendo nuevos públicos. Otra mirada se centra en la retención de socios, que en épocas de crisis cobra mayor importancia.

Dado el avance de la tecnología y la decisión de las organizaciones de contar con equipos de análisis, existe la posibilidad de evaluar, diseñar e implementar modelos o sistemas que puedan alertar previo a que los clientes pidan la baja del sistema, permitiendo así tomar medidas adecuadas antes de dicho momento.

En este marco, se plantea abordar el tema de retención de afiliados mediante técnicas de aprendizaje supervisado, las cuales permiten generar un modelo de clasificación que alerte la posibilidad de que un afiliado sea propenso a abandonar la empresa.

Esta propuesta abarca el estudio del comportamiento de todos los socios que estuvieron activos en algún momento en el período comprendido entre el 1 de enero de 2017 y el 31 de diciembre de 2018. Para este conjunto de personas se evalúan variables sociodemográficas, de uso del servicio de salud, asociadas al tipo de servicio contratado y condiciones comerciales elegidas (descuentos, promociones, medios de pago), al cumplimiento del pago, a los reclamos y soluciones brindadas.

## OBJETIVO

El objetivo principal de esta práctica profesional es construir un modelo estadístico que permita predecir la posibilidad de que un afiliado de Aca Salud decida darse de baja de la cobertura.

## MATERIAL DISPONIBLE

Las variables consideradas como posibles predictoras de las bajas de los afiliados surgen de múltiples reuniones de discusión llevadas a cabo entre los integrantes de la gerencia de Planeamiento Económico, Financiero y de la Operación. Ellas son: Tipo Afiliado

(Titular o Familiar), Sexo, Tipo de Afiliación (Voluntario u Obligatorio), Credencial (Integral, Superior, Selecta u Otras), Grupo Etario (De 0 a 25, De 26 a 45, De 46 a 60 o Más de 60), Antigüedad (Menos de 6 meses, De 6 meses a 1 año, De 1 a 3 o Más de 3 años), Estado Civil (Casado/a o Concubinato, Soltero/a o Viudo/a, Separado/a o Divorciado/a), entre otras.

Se decidió incluir a los afiliados que hayan estado activos, en algún momento, en los años 2017-2018, registrando su estado en esta última fecha. Se excluyeron los socios cuya baja del servicio ocurrió por la disolución del vínculo contractual entre la empresa que los nuclea y Aca Salud. También se excluyeron todos los empleados de Aca Salud que hoy tienen esta cobertura como beneficio, dado que la baja, en general, depende de la desvinculación del empleado y no responde al patrón de bajas del resto de los socios.

Durante la tarea de depuración se encontraron casos que presentaban inconsistencias y/o datos faltantes. Las inconsistencias se evaluaron una a una. Respecto a los datos faltantes, se completaron en su totalidad con información recolectada de otras fuentes (declaraciones juradas, etc.). Finalmente, luego de la depuración y enriquecimiento de la base de datos, la cantidad de socios que participan de este estudio es 161.122.

La variable de interés en este estudio es el Estado de cada afiliado al 31/12/2018 cuya categorización puede tomar dos valores: “Alta” (aquellos socios que están activos y cuentan con cobertura de salud) o “Baja” (socios que dieron de baja la cobertura). Entre los 161.122 afiliados que se tuvieron en cuenta para el análisis, el 21,8% estaban dados de baja al momento de finalización del estudio.

Con relación al sexo, el 50,8% de los afiliados en el universo bajo estudio son mujeres. Se observó que el porcentaje de bajas en los hombres es similar al de las mujeres y hay menos bajas en los afiliados de mayor edad. Además, si se analiza la permanencia de los afiliados en la empresa, el porcentaje más bajo de socios que ya no tienen la cobertura se presenta en aquellos que tienen más de 3 años de antigüedad.

## MÉTODOS DE APRENDIZAJE SUPERVISADO

Las técnicas de aprendizaje supervisado suponen que para cada observación del vector de variables explicativas existe una respuesta asociada. Mediante estos métodos se pretende ajustar un modelo que relacione la respuesta con los predictores de manera que sea posible entender mejor la asociación entre estas variables y/o predecir correctamente la respuesta para observaciones futuras (James, Witten, Hastie, & Tibshirani, 2017). Las técnicas de aprendizaje supervisado aplicadas en esta práctica profesional son: regresión logística y métodos basados en árboles tales como árboles de clasificación, *bagging* y *random forests*.

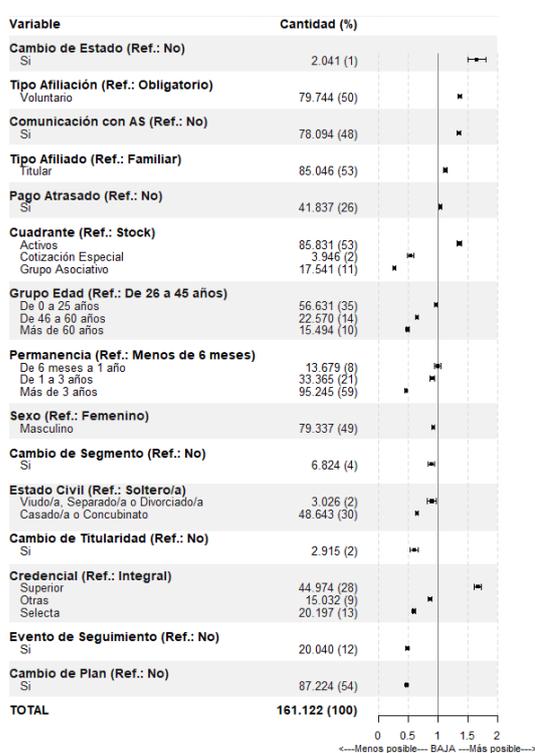
### Regresión Logística

Para cumplir con el objetivo de predecir la probabilidad de que un afiliado de Aca Salud decida darse de baja de la cobertura, se utiliza un modelo de regresión logística. En este

caso, la variable respuesta es el Estado del afiliado al momento de finalización del estudio y puede tomar las categorías “Alta” o “Baja”.

La estrategia originalmente planteada para la construcción del modelo multivariado comienza haciendo, a modo de screening una regresión logística univariada para cada una de las covariables. Se observa que todas las variables resultan significativas para la explicación del Estado debido a la gran cantidad de observaciones con las que se cuenta en la base de datos. Se analizan las razones de *odds* estimadas para cada uno de los modelos univariados y sus correspondientes intervalos de confianza para poder visualizar de manera más simple y clara cuáles son las covariables con mayor influencia en la respuesta (Figura 1).

Figura 1. Razones de *odds* calculadas a partir de los coeficientes de los modelos univariados y sus intervalos de confianza del 95%



Luego se estima este modelo multivariado incluyendo todas las variables que habían sido tenidas en cuenta en la etapa anterior.

La forma recomendable para trabajar con técnicas de aprendizaje supervisado consiste en dividir el conjunto de datos en una “base de entrenamiento” y una “base de prueba” ya que usar la misma base para ambas etapas podría arrojar resultados sesgados. Con la “base de entrenamiento” se construye el modelo y su predicción se realiza sobre los individuos de la “base de prueba” (James, Witten, Hastie, & Tibshirani, 2017). En este caso, se extrae una muestra al azar del 75% de los individuos para la muestra de entrenamiento y el 25% restante para la de prueba.

A partir del modelo de regresión logística se predicen las probabilidades de que un afiliado solicite la baja de la cobertura y luego se categorizan dichos valores tomando

como punto de corte el equilibrio óptimo entre especificidad y sensibilidad para poder comparar la respuesta predicha con el valor real observado. Por ello, se decide que toda probabilidad predicha mayor a 0,5 se considere como “Baja” y las menores como “Alta”. Este modelo clasifica correctamente al 78,44% de las observaciones de la muestra de prueba. Sin embargo, la sensibilidad del modelo es muy baja ya que sólo clasifica como “Baja” a 22 de las 8.693 bajas observadas realmente. Es decir que el modelo arroja una gran cantidad de falsos negativos y tiene una sensibilidad de 0,25%.

Para intentar mejorar la capacidad de clasificación correcta del modelo se decide agregarle interacciones de segundo grado. Sin embargo, no resulta posible incorporar todas las interacciones dobles debido a problemas de multicolinealidad.

A partir de este nuevo modelo, se clasifica correctamente al 79,63% de los afiliados de la muestra de prueba y la sensibilidad es del 14,53%. Claramente, la inclusión de interacciones mejora la capacidad de predicción del modelo. No obstante, la cantidad de parámetros de dicho modelo es muy grande y resulta compleja su interpretación.

Con el fin de mejorar aún más la sensibilidad se probaron modelos sobre conjuntos de datos reducidos (socios voluntarios, socios titulares, etc.). Sin embargo, no se obtuvieron mejoras en los resultados.

En conclusión, el modelo de regresión logística que mejor clasifica a la variable “Estado” es aquel en el que se incluyen las 15 variables explicativas junto con 104 interacciones dobles. Debido a la falta de parsimonia, interpretabilidad y ajuste, estos modelos son descartados.

## **Métodos basados en árboles**

Los métodos para clasificación basados en árboles estratifican o segmentan el espacio de los predictores en un número simple de regiones y, para obtener las predicciones, se suele usar la categoría más común (moda) de las observaciones en cada región definida. Los árboles de clasificación son simples y fáciles de interpretar, pero pueden no resultar lo suficientemente competitivos frente a otros métodos de aprendizaje supervisado en cuanto a la precisión de predicción. Los métodos de *bagging* y *random forests* producen múltiples árboles que se combinan para mejorar la predicción, a expensas de una interpretación más complicada.

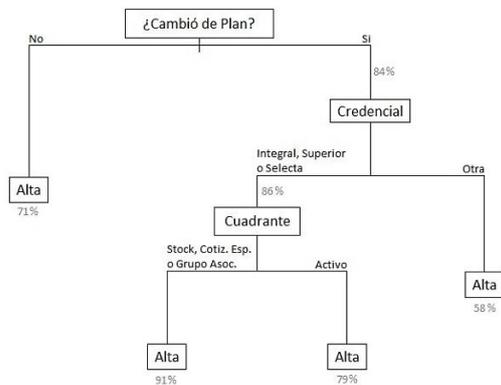
### *Árboles de clasificación*

Para generar el árbol de clasificación se aplica la división binaria recursiva. Este proceso comienza en el punto más alto del árbol (donde todas las observaciones pertenecen a una misma región), y de manera sucesiva se divide el espacio del predictor, donde cada división genera dos nuevas ramas y se categoriza en base a la respuesta más frecuente que se presenta en la región definida para esa observación. Un criterio para estas divisiones es el Índice de Gini que es una medida de la variancia total en el conjunto de las  $K$  clases, o de la pureza de los nodos.

Para construir el árbol de clasificación para la variable respuesta “Estado” se trabaja con

la misma muestra de entrenamiento y las mismas covariables utilizadas en el caso de regresión logística. El árbol de clasificación se calculó con el paquete *tree* en R. Lamentablemente, aun podando el árbol, el método tiene como predicción resultante en todas las ramas la categoría “Alta” de la respuesta “Estado” (Figura 2).

Figura 2. Árbol de clasificación para la muestra de entrenamiento



Esto se debe a que el porcentaje de afiliados activos en cada rama es mayor al de socios dados de baja. Es por ello que se obtiene un porcentaje de clasificación correcta del 78,09%, correspondiente a todos los afiliados que son “Altas”. En este caso, la sensibilidad del método es del 0%, ya que clasifica mal a todos los socios que habían solicitado la baja de la cobertura (Figura 2).

### Bagging

Los árboles de clasificación “sufren” de alta variancia, lo que significa que, si se dividen al azar los datos de entrenamiento en dos grupos y se ajusta un árbol de clasificación a cada mitad, los resultados obtenidos podrían ser bastante diferentes. Por el contrario, un procedimiento con baja variancia da resultados parecidos aun aplicándose sobre *sets* de datos distintos. El método de *bagging* es un procedimiento utilizado para reducir la variancia de un método de aprendizaje estadístico.

En un problema de clasificación, se construyen  $B$  árboles a partir de pseudo-muestras generadas por *bootstrapping* y, dada una observación de prueba, se puede obtener la clase predicha por cada uno de ellos y escoger como predicción final para dicha observación la clase más común de entre las  $B$  predicciones.

El número de árboles ( $B$ ) no es un parámetro crítico a la hora de aplicar *bagging*. Ajustar un gran número de árboles no aumenta el riesgo de sobreajuste, por lo que se usa un número lo suficientemente alto como para alcanzar la estabilización en la reducción del error de la prueba.

Una manera directa de estimar dicho error es la estimación del error *out-of-bag*. En promedio, cada árbol generado por *bootstrapping* usa en torno a  $\frac{2}{3}$  de las observaciones.

El  $\frac{1}{3}$  restante de observaciones no usadas para ajustar cada árbol se conocen como *out-of-bag* (OOB). Por lo tanto, se puede obtener la predicción para la  $i$ -ésima observación

de prueba usando cada uno de los árboles en los cuales dicha observación sea OOB. Esto resulta en  $\frac{B}{3}$  predicciones para cada observación, por lo que, para obtener un solo valor, se escoge la clase mayoritaria. La estimación del error del test se corresponde con el error de clasificación.

Nuevamente, se utilizó la misma muestra de entrenamiento para aplicar el método *bagging* para datos de clasificación con el paquete “*randomForest*” de R. Se ajustaron 500 árboles para los cuales se usaron las 15 variables categóricas explicativas utilizadas en los métodos anteriores. El error OOB estimado fue del 20,44%.

Luego, se utilizó el modelo obtenido para la predicción de la respuesta “Estado” en la muestra de prueba. El porcentaje de clasificación correcta es del 79,96% y la sensibilidad es del 17,8%. Es decir que, con este método, se observan mejores resultados que con regresión logística sin depender de un modelo complicado por tantas interacciones.

Sin embargo, este método mejora la clasificación a expensas de la interpretabilidad ya que no es posible contar con la expresión del modelo ni con la imagen de las ramas como en el árbol de clasificación. Aun así, pueden determinarse cuáles son las variables más importantes a partir de la disminución media en el índice de Gini. La variable más importante para el método *bagging* fue la Credencial, siendo la credencial “Superior” la más relacionada a la baja.

### *Random Forests*

Bajo el supuesto de que se cuenta con un conjunto de datos que tiene un predictor muy importante o influyente que se destaca sobre el resto, todos o casi todos los árboles generados por *bagging* usarán este predictor en la primera ramificación, por lo que acabarán siendo similares unos a otros, y las predicciones entre ellos estarán altamente correlacionadas. En este escenario, la aplicación de *bagging* promediando valores correlacionados no consigue una reducción sustancial de la variancia con respecto a un solo árbol.

El método de *random forests* proporciona una mejora a los árboles combinados por *bagging* en cuanto a que elimina la correlación, teniendo en cuenta sólo un subgrupo de predictores en cada división. Al igual que en el *bagging*, se construye un número de árboles de clasificación a partir de pseudo-muestras generadas por *bootstrapping*. Esta vez se escoge, entre los  $p$  predictores, una muestra aleatoria de tamaño  $s$  como candidatos antes de cada división. En el caso de que se use el método para la clasificación de la respuesta,  $s \approx \sqrt{p}$ . Sólo se aplica la división a uno de los  $s$  predictores. Esto hace que, en promedio,  $\frac{(p-s)}{p}$  divisiones no tengan en cuenta el predictor más influyente, dando más oportunidades al resto.

Para este método se usa también el paquete “*randomForest*” de R. Se ajustan 500 árboles para los cuales se prueban en cada separación 4 de las 15 variables explicativas aleatorias diferentes. El error OOB estimado resulta del 19,88%. El porcentaje de clasificación correcta del método es del 80,36% y su sensibilidad del 14,76%.

Si bien el porcentaje de clasificación es el más alto de todas las técnicas que se probaron,

su sensibilidad, que es un valor que interesa para el cumplimiento del objetivo, no superó a la del método *bagging*.

## DISCUSIÓN

En este trabajo se detallan las actividades realizadas en la empresa de medicina prepaga Aca Salud. Cabe destacar la posibilidad que brindó la empresa para realizar este trabajo con la ayuda de sus colaboradores. La Práctica Profesional habilitó un espacio de aprendizaje mediante el trabajo interdisciplinario y la aplicación de temas aprendidos durante la carrera Licenciatura en Estadística derivando en resultados que fueron presentados a las gerencias interesadas y, además, en la Jornada de Jóvenes Investigadores que fue llevada a cabo en la Facultad de Ciencias Económicas y Estadística de la ciudad de Rosario el 5 de junio de 2019.

Con el fin de cumplir con el objetivo se llevaron a cabo todas las tareas específicas previstas tales como la construcción y depuración de la base de datos, el análisis descriptivo de variables, la construcción y comparación de modelos para elegir aquel que arrojara mejores resultados en la predicción de la respuesta y la presentación y discusión de los resultados preliminares en una jornada afín a los temas propuestos por el trabajo.

Desafortunadamente los resultados obtenidos con las distintas técnicas estadísticas no son suficientes para dar respuesta al objetivo del estudio y, por ende, a la problemática de la empresa. Si bien se detectaron altos porcentajes de clasificación correcta de la variable respuesta con la mayoría de las técnicas, ninguna arrojó un resultado de sensibilidad aceptable. El porcentaje más alto de buena predicción de la categoría “Baja” se dio con la técnica *bagging*.

Estos resultados llevan a reflexionar acerca de que las variables que se consideraron para este estudio no son suficientes para detectar a aquellos afiliados que dieron de baja la cobertura de Aca Salud.

El principal aporte de este trabajo se centra en recomendar el registro y posterior inclusión de nuevas variables en la base de datos que permitan describir características de los socios con las que actualmente no se cuenta. Por ejemplo, podría intentar determinarse el nivel socioeconómico de cada individuo o su situación laboral ya que, en base a encuestas que se han realizado en la empresa, se conoce que el motivo de baja más frecuente es “Por cuestiones económicas”, sin embargo, en la actualidad esta información no está captada en forma completa y rutinaria.

Podría sospecharse además que, si bien se realizó una depuración de la misma, existían variables sobre las que surgían dudas o provocaban una falta de confianza en la información.

Estos resultados deben ayudar a generar conciencia dentro de la empresa. Los datos recabados son el capital más valioso de una empresa de servicio y por lo tanto se debe abogar por mejorar cada día su calidad.

Si bien estos resultados no son suficientes para abordar la problemática desde la

perspectiva de retención de bajas, tienen utilidad para poder enfocar la fuerza de venta sobre los perfiles de afiliados cuyo perfil coincide con los que tienen mayor posibilidad de seguir activos en la cobertura, es decir, los socios más fieles.

## BIBLIOGRAFÍA

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc.
- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (25 de Marzo de 2018). *Breiman and Cutler's Random Forests for Classification and Regression*. Recuperado el 01 de Junio de 2019, de <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Cramer, H. (1962). *Mathematical Methods of Statistics*. Asia Publishing House. Recuperado el 01 de Junio de 2019, de [http://fma.if.usp.br/~aalves/FPE/hc\\_probability.pdf](http://fma.if.usp.br/~aalves/FPE/hc_probability.pdf)
- Goodman, L., & Kruskal, W. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, 49(268), 732-764.
- Hosmer, D., & Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley & Sons, Inc.
- INDEC. (2010). *Censo Nacional de Población y Vivienda*. Argentina. Recuperado el 01 de Junio de 2019, de [https://www.indec.gob.ar/ftp/cuadros/poblacion/censo2010\\_tomo1.pdf](https://www.indec.gob.ar/ftp/cuadros/poblacion/censo2010_tomo1.pdf)
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning with Applications in R*. Springer.
- R project. (26 de Abril de 2019). *Fitting Generalized Linear Models*. Recuperado el 01 de Junio de 2019, de <https://www.rdocumentation.org/packages/stats/versions/3.6.1/topics/glm>
- Ripley, B. (26 de Abril de 2019). *Classification and Regression Trees*. Recuperado el 01 de Junio de 2019, de <https://cran.r-project.org/web/packages/tree/tree.pdf>

# ESTIMACIÓN EN DISEÑOS MUESTRALES COMPLEJOS EN PRESENCIA DE NO RESPUESTA

LIC. TOMÁS CAPRETTO

Director: MG. PABLO GONZALO MARÍ

Los estudios por muestreo están expuestos a diferentes tipos de errores, muestrales y no muestrales. Uno de los tipos de errores no muestrales es la no respuesta. Este trabajo presenta diferentes estrategias para mitigar el sesgo que puede introducir la no respuesta. Estas modificaciones dificultan la obtención de expresiones para la variancia. Se propone al estimador de variancia Bootstrap presentado por Rao, Wu y Yue (1992). Se realiza un estudio por simulación donde se busca imitar un diseño muestral complejo a gran escala con presencia de no respuesta, y se evalúa empíricamente el desempeño de los estimadores puntuales y de variancia.



## INTRODUCCIÓN

Uno de los métodos de obtención de datos más utilizados es el estudio por muestreo. En vez de indagar exhaustivamente sobre el conjunto finito de unidades que se desea estudiar, se hace una investigación utilizando una parte de éste.

Los estudios por muestreo están expuestos a diferentes tipos de errores, muestrales y no muestrales. El error muestral se produce por estimar una característica de interés de una población con sólo una parte de ésta. Al resto de los errores se los denomina no muestrales, entre los cuales el de no respuesta se presenta en la mayoría de los operativos. El error de no respuesta ocurre cuando se obtiene respuesta de solamente un subconjunto de la muestra que difiere del conjunto de unidades que no responden en características de interés de la encuesta. Consecuentemente, este error introduce sesgo en las estimaciones. Una solución viable y popular en la práctica es realizar algún tratamiento para disminuir el impacto de la no respuesta en el sesgo de los estimadores utilizados. Algunos métodos se basan en el ajuste de los pesos muestrales de los que responden.

Por otro lado, la estimación de variancia del estimador es una medida del error muestral y es necesaria por dos motivos: indicar la precisión del estimador y calcular intervalos de confianza. Las expresiones para la variancia no sólo dependen del diseño muestral, sino que también son influenciadas por el método utilizado para corregir la no respuesta.

Una alternativa plausible es utilizar algún método de replicación. Su principal beneficio es que se utiliza una única fórmula de estimación de variancia para cualquier estimador.

En particular, Rao, Wu y Yue (1992) proponen un método de estimación de variancia Bootstrap que es consistente incluso para estadísticos no suaves (la función que los define posee derivada primera continua), como la mediana u otro cuantil. Lamentablemente, no hay muchos resultados teóricos publicados al respecto de sus propiedades en diseños muestrales complejos.

Considerando un diseño muestral complejo y diferentes escenarios de no respuesta, este trabajo busca evaluar empíricamente las propiedades de diversas estrategias de ajuste de pesos muestrales, con principal foco en la capacidad para reducir el sesgo por no respuesta. Las estimaciones de variancia se realizan mediante el método propuesto por Rao, Wu y Yue (1992) a fin de conocer su desempeño en un escenario que se asemeja a los estudios por muestreo realizados por los organismos oficiales de estadística.

Para realizar las simulaciones necesarias, se utiliza una población sintética (población cuyas unidades fueron simuladas de forma tal que los agregados a diferentes niveles sociodemográficos se asemejan a la población verdadera) de viviendas del Estado de Texas correspondiente al año 2010. La misma es uno de los componentes de una población sintética elaborada para todo Estados Unidos por *Research Triangle Institute (RTI) International* que incluye personas, escuelas y diversas instituciones (Wheaton, 2014).

## CONCEPTOS BÁSICOS

### Muestreo en poblaciones finitas

El muestreo en poblaciones finitas considera un conjunto de técnicas dentro de la Estadística que tiene como objetivo obtener estimaciones de ciertos parámetros de una población, los cuales son fijos y desconocidos. El muestreo consiste en la selección de una parte de la población de forma tal que permite obtener resultados para la misma. El muestreo es probabilístico cuando el método utilizado para seleccionar la muestra asigna a cada unidad de la población de interés

una probabilidad no nula y conocida de ser seleccionada. Esto posibilita inferir sobre medidas de interés en la población utilizando reglas de la inferencia estadística para obtener estimaciones puntuales y cuantificar su incertidumbre.

Sea el conjunto  $U = \{u_1, \dots, u_k, \dots, u_N\}$  de tamaño  $N$ , conocido, la población de interés. Por simplicidad se la suele notar  $U = \{1, \dots, k, \dots, N\}$ . La característica de interés  $y$  presenta el valor  $y_k$  para cada unidad  $k$  de la población. El objetivo es estudiar una función de los  $y_k$ ,  $\theta = f(y_1, y_2, \dots, y_N)$ .

como por ejemplo el total  $t_y = \sum_{k \in U} y_k$  o la media  $\bar{y}_U = \frac{1}{N} \sum_{k \in U} y_k$ .

Una muestra es un subconjunto de la población  $s \subset U$ . Un diseño muestral es una distribución de probabilidad sobre todas las muestras posibles que se pueden obtener de  $U$ .

La probabilidad de que una unidad  $k$  pertenezca a una muestra se denomina probabilidad de inclusión de primer orden de la unidad  $k$  y se obtiene como la suma de las probabilidades de todas las muestras a la que  $k$  pertenece:

$$\pi_k = P(\text{unidad } k \in s) = \sum_{s \ni k} p(s),$$

donde  $s \ni k$  se lee como "las muestras  $s$  que contienen al elemento  $k$ ".

El recíproco de  $\pi_k$  es el peso muestral de la unidad  $k$ ,  $d_k = 1/\pi_k$ .

La probabilidad de que dos unidades  $k$  y  $l$  pertenezcan a una muestra se denomina probabilidad de inclusión de segundo orden de las unidades  $k$  y  $l$ . La misma se obtiene mediante la suma de las probabilidades de todas las muestras a las que  $k$  y  $l$  pertenecen:

$$\pi_{kl} = P(\text{unidades } k \text{ y } l \in s) = \sum_{s \ni k \& l} p(s).$$

Éstas permiten definir los estimadores clásicos del muestreo en poblaciones finitas.

### Estimadores clásicos del muestreo en poblaciones finitas

Un estimador es una función de la muestra que se espera produzca valores cercanos al parámetro de interés en la mayoría de las muestras.

El estimador de Horvitz-Thompson para el promedio poblacional es:

$$\hat{y}_\pi = \frac{1}{N} \sum_{k \in s} \frac{y_k}{\pi_k}.$$

Es un estimador lineal y se puede encontrar fácilmente una expresión para su variancia y su correspondiente estimación. Una alternativa que no requiere conocer el tamaño poblacional  $N$  es el estimador de Hájek (1971):

$$\hat{y}_{HJ} = \frac{\sum_{k \in s} y_k / \pi_k}{\sum_{k \in s} 1 / \pi_k},$$

que es no lineal por ser una razón entre dos estimadores. Se lo puede aproximar por series de Taylor y obtener una expresión de la variancia para el estimador linealizado.

En general los Institutos de Estadística utilizan al estimador de Hájek por sobre el de Horvitz-Thompson. Uno de los motivos es que el estimador de Hájek manifiesta cierta insensibilidad ante muestras poco probables.

## DISEÑOS MUESTRALES COMPLEJOS

En el contexto de las estadísticas oficiales los parámetros de interés se corresponden con poblaciones distribuidas en un territorio geográfico muy extenso, por ejemplo, todo un país. El estudio de estos parámetros es posible mediante la utilización de diseños que seleccionan la muestra de una forma más sofisticada que un muestreo simple al azar.

Un diseño muestral complejo es aquel que presenta las siguientes características:

1. Distintas probabilidades de selección;
2. Múltiples etapas de muestreo;
3. Estratificación de las unidades de al menos una etapa;
4. Selección de conglomerados en etapas iniciales.

## EL PROBLEMA DE LA NO RESPUESTA

Cuando se obtienen cuestionarios no completos en su totalidad, existe un problema de no respuesta. En primer lugar, ésta disminuye el tamaño de muestra efectivo, lo que incrementa el error de muestreo y limita la capacidad de generalizar conclusiones. En segundo lugar, la no respuesta puede introducir sesgo en las estimaciones de interés.

### Sesgo por no respuesta

Sea  $I_k$  una variable dicotómica que indica la pertenencia de la unidad  $k$  en la muestra. Sea  $R_k$  una variable dicotómica que indica si la unidad  $k$  responde, dado que está en la muestra. Luego  $P(I_k = 1) = \pi_k$  y la probabilidad condicional de responder es  $P(R_k = 1 | I_k = 1) = \phi_k$ . Esta cantidad también se conoce como *propensity score* de la unidad  $k$ .

El sesgo por no respuesta del estimador de Hájek para la media se puede expresar

$$B(\hat{y}_{HJ}) \cong CV_\phi \sigma_y \rho_{\phi,y},$$

donde:

- $CV_\phi$  es el coeficiente de variación de la propensión a responder.
- $\sigma_y$  es el desvío estándar poblacional de la variable en estudio.
- $\rho_{\phi,y}$  es la correlación poblacional entre la propensión a responder y la variable en estudio.

$\hat{y}_{HJ}$  es aproximadamente insesgado cuando se cumple al menos una de las siguientes condiciones:

- i. La correlación entre la variable de interés y la probabilidad de responder es nula;
- ii. Todas las unidades tienen la misma probabilidad de responder.

Por otro lado, cuanto mayor sea la variabilidad de la probabilidad de responder y su correlación con la característica de interés, mayor el sesgo por no respuesta.

## Ajuste por no respuesta

En la literatura se encuentran diversas sugerencias para tratar la no respuesta mediante un incremento de la tasa de respuesta: mayor difusión de los operativos, más intentos de contactos o personalizar el enfoque de la encuesta. Sin embargo, los Institutos de Estadística vienen reportando disminución de la tasa de respuesta.

Una solución viable y popular en la práctica es realizar algún tratamiento que permita disminuir el sesgo por no respuesta. Como lo sugiere la palabra ajuste, se realizan modificaciones a un procedimiento de estimación original o ideal que hubiera sido utilizado en el caso de que todas las unidades respondiesen.

Por un lado, se tienen los métodos de imputación. Estos reemplazan a los datos faltantes con "valores artificiales". Por otro lado, se tiene el ajuste de pesos muestrales. Estos incrementan los pesos muestrales de los que responden de forma tal que se compense la falta de respuesta cuando se realizan las estimaciones.

A su vez, dentro del ajuste de pesos muestrales se encuentran dos grandes grupos: el enfoque de *propensity score* y el enfoque de calibración. El primero busca disminuir el sesgo por no respuesta a través de la estimación directa de la probabilidad de responder y lo utiliza para escalar los pesos muestrales. El segundo ajusta los pesos muestrales de los que responden de forma tal que las estimaciones de variables auxiliares coincidan con sus totales poblacionales conocidos.

### Ajuste mediante la propensión a responder

Se asume que  $\phi_k > 0 \forall k \in U$  y que las unidades responden independientemente. Mediante algún modelo paramétrico o no paramétrico se estima la probabilidad de responder para cada unidad,  $\hat{\phi}_k$ . Luego se obtiene un conjunto de pesos  $\{\tilde{w}_k\}_{k \in S_r}$  donde  $\tilde{w}_k = d_k / \hat{\phi}_k$ .

En la práctica se suelen crear clases donde se agrupan a los que responden y se aplica un factor de ajuste común. Independientemente del método elegido, una forma posible de calcular el factor de ajuste para la clase  $c$  es:

$$\hat{\phi}_k^{-1} = a_c = \frac{\sum_{k \in S_c} d_k}{\sum_{k \in S_{r,c}} d_k},$$

la razón entre la suma de los pesos de las unidades de la clase y la suma de los pesos de las unidades de la clase que responden. Luego se realizan las estimaciones con  $\tilde{w}_k$  en vez de  $d_k$ .

A medida que las estimaciones de la probabilidad de responder se asemejan más a su valor verdadero, el sesgo por no respuesta decrece. En el caso hipotético que coincidan, se podrían obtener estimadores insesgados.

### Ajuste mediante calibración

El método de calibración constituye un enfoque general y sistemático para la utilización de información auxiliar. Una gran variedad de estimadores utilizados para ajustar por no respuesta están comprendidos bajo este enfoque.

Sea  $\mathbf{x}_k$  el vector de variables auxiliares para la unidad  $k$  y sea  $\mathbf{t}_x = \sum_{k \in U} \mathbf{x}_k$ , vector de totales poblacionales conocidos. Un conjunto de pesos calibrados  $\{w_k\}_{k \in S}$  es aquel que satisface la ecuación de calibración  $\sum_{k \in S} w_k \mathbf{x}_k = \mathbf{t}_x$ .

La elección del conjunto de pesos calibrados se hace de forma tal que una medida de distancia entre los vectores  $(d_1, d_2, \dots, d_n)'$  y  $(w_1, w_2, \dots, w_n)'$  sea mínima, sujeto a la ecuación de calibración. Esto resulta en:

$$w_k = d_k F(\mathbf{x}'_k \boldsymbol{\lambda}),$$

donde  $F$  depende de la función de distancia utilizada y  $\boldsymbol{\lambda}$  es un vector de multiplicadores de Lagrange.

Uno de los motivos por los cuales se suele utilizar este enfoque es que ayuda a cumplimentar el doble objetivo de reducir el error muestral y el error por no respuesta, siempre y cuando se cuente con buena información auxiliar. Además, puede ser aplicado para cualquiera de los diseños muestrales utilizados frecuentemente y para cualquier vector auxiliar de información.

### Ajuste en uno y dos pasos

Se conoce como ajuste en un paso a aquel ajuste donde los pesos son solamente calibrados. Los pesos que se obtienen se pueden denominar  $w_k^{C,1}$ . Por otro lado, el ajuste a dos pasos consiste en primero ajustar a los pesos mediante una estimación de la propensión a responder, y luego calibrar estos pesos. Se los suele denotar  $w_k^{C,2}$ .

El paso del ajuste por la propensión a responder busca reducir el sesgo por no respuesta, y la calibración busca consistencia con totales conocidos y reducir la variancia. La capacidad que tenga este estimador para cumplir los objetivos planteados depende del método utilizado para estimar la propensión a responder y de cuán relacionada esté  $y_k$  con  $\mathbf{x}_k$ .

### ESTIMACIÓN DE VARIANCIA

El problema que presentan los ajustes de pesos muestrales mencionados es que producen estimadores no lineales. Una alternativa es utilizar algún método de replicación. La idea es obtener una serie de submuestras (réplicas), y luego utilizar cada una de ellas para estimar al parámetro con el mismo procedimiento que en la muestra completa. La variancia es estimada a partir de la variabilidad entre las estimaciones en cada replicación.

Uno de los métodos de replicación más populares es *Bootstrap*. Éste fue presentado por Efron (1979) en el contexto de la estadística clásica, donde las observaciones se asumen independientes e igualmente distribuidas. Rao, Wu y Yue (1992) proponen un método de estimación de variancia *Bootstrap* para diseños estratificados multietápicos que realiza un escalado en los pesos muestrales y aplica tanto para estimadores suaves como no suaves.

La gran ventaja de este método, y de los métodos de replicación en general, es que permiten reflejar implícitamente los efectos en la variancia de los ajustes por no respuesta y calibración. El escalado de los pesos propuesto por Rao, Wu y Yue se corresponde con un diseño donde el muestreo de unidades primarias es realizado con reemplazo. La utilidad de este método en la práctica se debe a que sobreestima la variancia verdadera para ciertos diseños muestrales.

## ESTUDIO POR SIMULACIÓN

### Características del estudio por simulación

Este trabajo propone un hipotético estudio por muestreo que tiene por objetivo estimar el ingreso medio de los hogares de un Estado. Se selecciona una muestra probabilística de 31300 viviendas mediante un diseño por conglomerados estratificado en tres etapas.

- i. **Etapa 1 (UPM):** seleccionar una muestra estratificada  $s_r$  de Unidades Primarias de Muestreo (en este caso aglomerados urbanos) con probabilidad proporcional a la cantidad de viviendas mediante el método sistemático de Madow (Madow, 1949). Aquellos aglomerados urbanos más importantes conforman un estrato de inclusión forzosa.
- ii. **Etapa 2 (USM):** para cada UPM  $\in s_r$  se selecciona una muestra  $s_{Ii}$  de áreas de muestreo mediante una selección simple al azar sin reemplazo.
- iii. **Etapa 3:** dentro de cada USM  $\in s_{Ii}$  se selecciona una muestra  $s_{iq}$  de viviendas mediante muestreo simple al azar sin reemplazo.

Tabla 1. Cantidad de unidades seleccionadas por etapa y por estrato de UPM

Etapa	Estrato de UPM			
	1	2	3	4
I	30	20	8	7
II	10	30	60	250
III	10	10	10	10
Total de viviendas	3000	6000	4800	17500

### Asignación de valores faltantes

Dada una muestra, se asigna un porcentaje de valores faltantes de forma que se emulen diferentes mecanismos de no respuesta.

- i. **Missing completely at random (MCAR):** se le asigna a cada unidad la misma probabilidad de responder, igual a la tasa de respuesta deseada.
- ii. **Missing at Random (MAR):** en este caso cada unidad muestreada tiene una probabilidad de responder proporcional a la cantidad de personas por hogar.
- iii. **Not missing at random (NMAR):** cada vivienda tiene una probabilidad de responder inversamente proporcional al ingreso del hogar, que es la variable de interés.

### Ajuste de pesos muestrales

En cada escenario de no respuesta se prueban diversos ajustes de pesos. Sea  $w_k^*$  el peso definitivo para la unidad  $k$ , con  $k \in s_r$ .

- i. Sin ajuste: se utilizan los pesos de diseño,  $w_k^* = d_k$ , y el estimador recibe el nombre  $\hat{y}_{HJR}$ .
- ii. Calibración: los pesos de diseño son calibrados según conteos poblacionales marginales de personas por intervalos de edad y sexo,  $w_k^* = d_k F(x_k' \lambda)$  y el estimador es  $\hat{y}_{CAL}$ .

- iii. Ajuste por no respuesta 1 + Calibración: se ajustan los pesos utilizando a la cantidad de personas por hogar como variable de clasificación y luego son calibrados de la misma forma que en (ii),  $w_k^* = d_k \hat{\phi}_k^{-1} F(x_k' \lambda)$  y el estimador es  $\hat{y}_{NR1+CAL}$ .
- iv. Ajuste por no respuesta 2 + Calibración: se ajustan los pesos utilizando al estrato de UPM como variable de clasificación y luego son calibrados de la misma forma que en (ii),  $w_k^* = d_k \hat{\phi}_k^{-1} F(x_k' \lambda)$  y el estimador es  $\hat{y}_{NR2+CAL}$ .

### Estimadores utilizados

El parámetro de interés es el ingreso medio por hogar,  $\bar{y}_U$ . se utiliza al estimador de Hájek calculado sobre el conjunto de unidades que responden y sus pesos vienen dados cada uno de los ajustes anteriormente planteados,

$$\hat{y}_{HJr} = \frac{\sum_{k \in S_r} w_k^* y_k}{\sum_{k \in S_r} w_k^*}.$$

En todos los escenarios, y para cada una de las muestras seleccionadas, se estima la variancia con *Bootstrap* según Rao, Wu y Yue (1992).

### Evaluación de los estimadores

En el estudio por simulación se obtienen 10.000 muestras y se usan  $B = 200$  réplicas en *Bootstrap*. Para la evaluación de los estimadores puntuales y de variancia se utilizan las siguientes medidas:

- i. Sesgo relativo del estimador puntual  $RB_{SIM}(\hat{y})$
- ii. Variancia simulada del estimador puntual  $V_{SIM}(\hat{y})$
- iii. Error cuadrático medio del estimador puntual  $ECM_{SIM}(\hat{y})$
- iv. Sesgo relativo del estimador de variancia  $RB_{SIM}[\hat{V}(\hat{y})]$
- v. Variancia simulada del estimador de variancia

### Resultados

A continuación, se presentan tablas resumen con los resultados de la simulación. En las mismas se contemplan dos escenarios para la severidad de la tasa de no respuesta, 20% y 40% respectivamente.

Tabla 2. Resumen de resultados de la simulación. Tasa de no respuesta = 20%

Mecanismo de No Respuesta	Estimador Puntual	$RB_{SIM}(\hat{y})$ $\times 10^2$	$V_{SIM}(\hat{y})$ $\times 10^{-6}$	$ECM_{SIM}(\hat{y})$ $\times 10^{-6}$	$RB_{SIM}[\hat{V}(\hat{y})]$ $\times 10^2$	$V_{SIM}[\hat{V}(\hat{y})]$ $\times 10^{-11}$
MCAR	$\hat{y}_{HJR}$	0.005%	2.213	2.213	31.37%	5.971
	$\hat{y}_{CAL}$	1.99%	1.902	3.735	26.23%	3.911
	$\hat{y}_{NR1+CAL}$	0.009%	1.766	1.766	27.85%	3.394
	$\hat{y}_{NR2+CAL}$	0.009%	1.770	1.766	27.70%	3.377
MAR	$\hat{y}_{HJR}$	1.69%	2.416	3.744	29.03%	6.865
	$\hat{y}_{CAL}$	3.88%	2.036	8.989	24.73%	4.423
	$\hat{y}_{NR1+CAL}$	0.01%	1.761	1.762	27.83%	3.327
	$\hat{y}_{NR2+CAL}$	1.96%	1.869	3.636	26.20%	3.585
NMAR	$\hat{y}_{HJR}$	-17.99%	1.188	15.088	30.91%	1.490
	$\hat{y}_{CAL}$	-15.54%	1.078	11.272	26.79%	1.123
	$\hat{y}_{NR1+CAL}$	-16.47%	1.030	12.675	27.94%	1.024
	$\hat{y}_{NR2+CAL}$	-17.07%	0.983	13.564	29.16%	0.935

Tabla 3. Resumen de resultados de la simulación. Tasa de no respuesta = 40%

Mecanismo de no Respuesta	Estimador Puntual	$RB_{SIM}(\hat{y})$ $\times 10^2$	$V_{SIM}(\hat{y})$ $\times 10^{-6}$	$ECM_{SIM}(\hat{y})$ $\times 10^{-6}$	$RB_{SIM}[\hat{V}(\hat{y})]$ $\times 10^2$	$V_{SIM}[\hat{V}(\hat{y})]$ $\times 10^{-11}$
MCAR	$\hat{y}_{HJR}$	0.004	2.300	2.301	30.68	6.443
	$\hat{y}_{CAL}$	4.36	2.139	10.927	24.54	4.972
	$\hat{y}_{NR1+CAL}$	0.04	1.846	1.847	27.04	3.685
	$\hat{y}_{NR2+CAL}$	0.04	1.850	1.851	27.11	3.675
MAR	$\hat{y}_{HJR}$	2.10	2.551	4.581	28.57	8.044
	$\hat{y}_{CAL}$	6.67	2.319	22.877	23.10	6.131
	$\hat{y}_{NR1+CAL}$	0.04	1.835	1.836	27.16	3.782
	$\hat{y}_{NR2+CAL}$	2.51	1.975	4.893	25.81	4.211
NMAR	$\hat{y}_{HJR}$	-26.92	0.896	33.594	29.05	0.777
	$\hat{y}_{CAL}$	-22.53	0.908	23.564	23.94	0.776
	$\hat{y}_{NR1+CAL}$	-24.61	0.827	28.083	26.29	0.625
	$\hat{y}_{NR2+CAL}$	-25.54	0.770	30.226	27.66	0.544

## CONCLUSIONES

### En términos generales

- Si la no respuesta es MCAR, no se introduce sesgo en las estimaciones puntuales, a excepción cuando se utiliza el estimador calibrado. Sin embargo, se incrementa el error

de muestreo. El sesgo en el estimador calibrado podría relacionarse con la función de distancia utilizada.

- Si la no respuesta es MAR y se conocen las variables relacionadas a la no respuesta, el sesgo es eliminado casi completamente. Este escenario es muy simplista e ideal.
- Si la no respuesta está exclusivamente relacionada a la variable en estudio, el sesgo es proporcional al porcentaje de respuesta y no se puede mitigar.

### En términos de la variancia

- El ajuste a un paso reduce la variancia del estimador puntual y del estimador de variancia.
- El ajuste a dos pasos conserva la reducción en el sesgo debido al ajuste mediante propensión a responder y la reducción en la variancia debida a la calibración.
- La disminución en la variancia es independiente del sesgo del estimador puntual.

### Respecto a *Bootstrap* como estimador de variancia

- Las estimaciones son muy conservadoras.
- Un posible motivo es que se utilizaron fracciones de muestreo no depreciables en la primera etapa.
- Existen alternativas que buscan utilizar las fracciones de muestreo en el proceso de estimación, pero no están muy estudiadas aún.

## BIBLIOGRAFÍA

- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.
- Hájek, J. (1971). Comment on An essay on the logical foundations of survey sampling by Basu, D. (R. a. Holt, Ed.) *Foundations of Statistical Inference*, 236.
- Haziza, D., & Lesage, É. (2016). A Discussion of Weighting Procedures for Unit Nonresponse. *Journal of Official Statistics*, 32(1), 129-145.
- Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663-685.
- Madow, W. G. (1949). On the Theory of Systematic Sampling, II. *The Annals of Mathematical Statistics*, 20(3), 333-354. doi:10.1214/aoms/1177729988
- Rao, J. N., Wu, C. F., & Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology*, 18(2), 209-217.
- Wheaton, W. D. (Mayo, 2014). 2010 U.S. Synthetic Population. RTI International. Recuperado de <https://www.epimodels.org/midas/Rpubsyntdata1.do>

# ESTRATEGIAS PARA LA CONSTRUCCIÓN DE DISEÑOS $2^{k-p}$ PLEGADOS DE ACUERDO A LOS CRITERIOS DE RESOLUCIÓN Y ABERRACIÓN

LIC. ANTONELLA N. MAZZEO

Directora: DRA. LUCÍA HERNÁNDEZ  
Codirector: DR. JOSÉ ALBERTO PAGURA

---

El uso de los diseños factoriales ha crecido debido al reconocimiento por expertos de que son esenciales para la mejora de la calidad de productos o servicios. Utilizar diseños factoriales fraccionarios ( $2^{k-p}$ ) es una alternativa cuando el número de pruebas a ensayar resulta muy elevado. En el caso de aplicarlos y requerir datos adicionales, la técnica tradicional es utilizar un diseño plegado. Sin embargo, esta metodología no es la más eficiente a la hora de lograr estimaciones de efectos que no se confundan entre sí. En los últimos años, diferentes autores han demostrado que se pueden obtener diseños más eficientes. Esta tesina se encarga de mostrar las nuevas estrategias propuestas por dichos autores para la construcción de diseños factoriales fraccionarios plegados de acuerdo a los criterios de resolución y aberración.



## INTRODUCCIÓN

En los últimos años, la aplicación de la metodología conocida como “Diseño de Experimentos” ha crecido notablemente debido al reconocimiento de que los experimentos diseñados son esenciales para la mejora de la calidad de productos y servicios. Los diseños factoriales son la mejor estrategia experimental para estudiar los efectos de múltiples factores sobre una característica o variable respuesta de interés. La eficiencia de estos diseños aumenta a medida que el número de factores considerados aumenta, y con ella el número de pruebas que deben ser realizadas. Si el número de factores a considerar es grande, una manera de reducir el número de pruebas es ensayar cada factor en dos niveles. Esto conduce a los llamados diseños  $2^k$  (k factores a 2 niveles).

En los casos en que el número de pruebas a ensayar es elevado, la mejor alternativa que se encuentra en la bibliografía es llevar a cabo una fracción de las pruebas requeridas por un diseño factorial completo, lo que conduce a un diseño factorial fraccionario  $2^{k-p}$  (donde  $p$  es la cantidad de efectos cuya estimación se está dispuesto a sacrificar), que es un diseño que consta de  $2^{k-p}$  tratamientos. Si se necesitan datos adicionales para poder resolver ambigüedades acerca de ciertos efectos, una estrategia comúnmente utilizada es llevar a cabo un nuevo experimento con otra fracción del diseño  $2^{k-p}$  inicial. Esta técnica se conoce como “diseño plegado”. La opción tradicionalmente utilizada consiste en ensayar la nueva fracción invirtiendo los niveles de todos los factores involucrados en el experimento.

En los últimos años algunos autores han mostrado que se pueden obtener diseños con mayor resolución o con la misma resolución, pero con mayor capacidad para separar efectos que el resultante del diseño plegado tradicional.

En esta tesina se presenta una reseña de las principales características de los diseños  $2^{k-p}$ , se describen las estrategias de plegado tradicionalmente utilizadas y se introducen las nuevas propuestas existentes en la bibliografía para construir el nuevo plan experimental. La conclusión general de este trabajo es que, para la mayoría de los diseños, existen mejores estrategias que la consistente en invertir los signos de todos los factores, las cuales logran diseños combinados con mayor resolución y mayor capacidad para separar efectos confundidos entre sí.

## METODOLOGÍA

### Diseños factoriales fraccionarios a 2 niveles

#### *Definición y construcción*

Los diseños factoriales son la mejor estrategia experimental para estudiar los efectos de múltiples factores sobre una característica o variable respuesta de interés. Este tipo de diseño presenta dos ventajas principales respecto de otros. La primera es que permiten detectar si existen interacciones entre los factores bajo estudio. La segunda es que son más eficientes, es decir, requieren menos pruebas para lograr la misma precisión en la estimación de efectos, que

si se llevara a cabo la estrategia de variar un factor a la vez separadamente. Los diseños factoriales consisten en ensayar al menos una prueba para cada una de las combinaciones de los niveles o variantes en los que se quieren estudiar los factores. La eficiencia de los diseños factoriales aumenta a medida que el número de factores considerados aumenta. Sin embargo, con el incremento del número de tratamientos aumenta el número de pruebas que deben ser realizadas. Si el número de factores a considerar es grande, una manera de reducir el número de pruebas es ensayar cada factor en dos niveles. Esto conduce a los llamados diseños  $2^k$ , que son los que requieren el menor número de pruebas para estudiar  $k$  factores, pero que, en el caso de factores cuantitativos, sólo permite detectar efectos lineales de los factores. Estos diseños tienen asociados técnicas y notaciones especiales. Los dos niveles para cada factor son denominados genéricamente “bajo” y “alto” y simbolizados con “-1” y “+1” respectivamente. Estos términos tienen significados evidentes en el caso de factores cuantitativos, pero son sólo etiquetas en el caso de factores cualitativos. El modelo estadístico asociado a un diseño  $2^k$  incluye  $k$  efectos principales,  $\binom{k}{2}$  interacciones dobles,  $\binom{k}{3}$  interacciones de tres factores, y así sucesivamente hasta llegar a la interacción de los  $k$  factores.

Una manera en la cual usualmente se presentan estos diseños es a través de una tabla de signos más y menos, o bien +1 y -1, que consiste en  $2^k$  filas, una por cada prueba, y  $k$  columnas, una por cada factor. El cuerpo de la tabla contiene +1 o -1 indicando el nivel al que es ensayado el factor de la columna en la prueba en cuestión. A modo de ejemplo, se considera el diseño  $2^3$  con tres factores  $A$ ,  $B$  y  $C$  cuya tabla de más y menos se presenta a continuación.

Tabla 1. Pruebas del diseño  $2^3$

Tratamiento	A	B	C
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

Una característica de los diseños  $2^k$  es que cada efecto principal y cada interacción tienen asociado un solo grado de libertad y es representado a través de un contraste. Cabe señalar que el término contraste refiere a una combinación lineal de las observaciones para la cual la

suma de los coeficientes es igual a cero, es decir,  $\sum_{i=1}^N c_i y_i$  donde  $\sum_{i=1}^N c_i = 0$ . Precisamente, la Tabla 1 presentada anteriormente provee los coeficientes asociados a los contrastes para el caso de los efectos principales. El efecto asociado a un factor está dado por la diferencia entre la respuesta media cuando el factor está a nivel “más” y la respuesta media cuando el factor está a nivel “menos”. En el ejemplo del diseño de la tabla 1, el efecto principal del factor C está dado por:

$$C = \bar{y}_+ - \bar{y}_- = \frac{y_5 + y_6 + y_7 + y_8}{4} - \frac{y_1 + y_2 + y_3 + y_4}{4}$$

Los contrastes que estiman las distintas interacciones pueden ser determinados de manera similar. Para ello, en primer lugar, se construyen las columnas asociadas a cada una de las interacciones realizando el producto elemento a elemento de las columnas de los factores intervinientes en la interacción en cuestión.

En las primeras etapas de una investigación, cuando el objetivo es determinar cuáles factores son importantes en una gran cantidad de candidatos, los diseños factoriales pueden requerir demasiadas pruebas, aun cuando los factores sean ensayados a sólo dos niveles, resultando en experimentos demasiado grandes y por lo tanto costosos. Un diseño **factorial fraccionario** es una modificación del diseño factorial completo que permite obtener información acerca de efectos principales e interacciones de orden bajo sin tener que realizar todas las pruebas que requiere el diseño factorial completo.

Un diseño factorial fraccionario  $2^{k-p}$ , o simplemente diseño  $2^{k-p}$ , es un diseño para  $k$  factores cada uno medido en dos niveles que utiliza  $2^{k-p}$  tratamientos. Esencialmente, este diseño consiste en dividir al conjunto de  $2^k$  tratamientos en  $2^p$  grupos y ensayar sólo uno de ellos. Lógicamente que, para que este enfoque sea efectivo, la fracción de las pruebas debe ser cuidadosamente elegida de modo que el plan experimental preserve alguna de las ventajas de los diseños factoriales completos.

Un diseño  $2^{k-p}$  se construye eligiendo  $p$  efectos cuya estimación se está dispuesto a sacrificar que, en general, son interacciones de orden alto. Estos efectos, llamados **generadores** o **palabras**, determinan grupos dividiendo a los  $2^k$  tratamientos según si sus signos son +1 o -1. Cada generador se escribe  $I = W$  o  $I = -W$  siendo  $W$  un efecto según corresponda a un bloque de +1 o de -1 respectivamente. La expresión  $I = W$  es llamada **relación de definición** del diseño.

### Confusión

Al ensayar un número menor de pruebas que las de un diseño  $2^k$ , se pierde parte de la información sobre algunos efectos. Más precisamente, se pierde la capacidad de estimar algunos efectos y la de distinguir o separar las estimaciones de grupos de los demás efectos. Dicho de otra manera, algunos efectos se estiman mediante exactamente el mismo contraste dando lugar a lo que se denomina confusión de efectos. En un diseño  $2^{k-p}$  la estimación de un efecto corresponde en realidad a  $2^p$  efectos, los cuales se dice que están *confundidos* entre sí o son *alias*. Si los efectos  $C$ ,  $AB$  y  $-ABCDE$  están confundidos entre sí, entonces el contraste que estima  $C$ , en realidad estima a la combinación  $C + AB - ABCDE$ . La confusión entre los efectos de un diseño particular da lugar a la denominada **estructura de alias**.

### Crterios para caracterizar y comparar fracciones factoriales $2^{k-p}$

Las fracciones factoriales son clasificadas de acuerdo con su **resolución** que es una medida que resume qué tipos de efectos están confundidos entre sí. Un diseño de resolución  $R$  es un

diseño en el que ninguna interacción de  $j$  factores está confundida con una interacción de menos de  $R - j$  factores.

En muchas situaciones, se encuentran diseños para igual número de factores con igual resolución, pero cuyas estructuras de alias difieren. En estos casos, el diseño preferible es aquel que tiene el menor número de palabras cortas, es decir la menor cantidad de interacciones de orden bajo, en la relación de definición. Este diseño es llamado diseño de **mínima aberración**. El número de factores en cada palabra de definición determina la longitud de la palabra en cuestión. La regla general es entonces máxima resolución y mínima aberración.

Una forma de resumir el número de palabras en la relación de definición de un diseño es a través del llamado **vector de longitud de palabras** dado por:

$$W = (w_3, \dots, w_k),$$

donde  $w_j$  es el número de palabras de longitud  $j$  siendo  $j = 3, \dots, k$  en la relación de definición del diseño.

### Diseños plegados

Si el análisis de los datos de un experimento de un diseño  $2^{k-p}$  arroja algunas conclusiones ambiguas, será necesario llevar a cabo un experimento adicional para resolverlas. El nuevo experimento deberá ser realizado en diferentes combinaciones de niveles de los factores respecto del experimento original. Esto proveerá información adicional que puede ser utilizada para separar los efectos confundidos. La técnica conocida como “*foldover*” o “plegado” consiste en realizar una nueva fracción de igual tamaño que la inicial cambiando los niveles de algunos o de todos los factores por aquellos niveles no utilizados en la fracción inicial (o dicho más simplemente, invirtiendo los niveles de los factores). Un plegado puede ser construido por dos motivos: para resolver ambigüedades entre algunos efectos en particular o para separar tantas estimaciones de efectos como sea posible. En el primer caso, la elección del plegado es bastante inmediata y consiste en invertir los niveles de los factores de interés particular solamente. En el segundo caso, la elección del plegado no es tan inmediata siendo la opción tradicionalmente utilizada la que consiste en invertir los niveles de todos los factores involucrados en el experimento. Esta estrategia es buena dado que, en muchos casos, el diseño combinado, es decir el diseño consistente de la fracción inicial y de la nueva fracción, tendrá mayor resolución. Sin embargo, en los últimos años algunos autores han mostrado que se pueden obtener diseños con mayor resolución o con la misma resolución, pero con mayor capacidad para separar efectos que el resultante del plegado tradicional.

En el año 1996, Montgomery y Runger publicaron un artículo que presenta dos reglas que pueden ser utilizadas para seleccionar y evaluar plegados para diseños  $2^{k-p}$ . La primera de ellas muestra que se puede describir un plegado en términos de la relación de definición completa de la fracción original, de la nueva fracción y del diseño combinado en lugar de en términos de los generadores. En su artículo, los autores utilizaron estas reglas para construir diseños plegados para algunos diseños de resolución *IV* particulares. Estos ejemplos muestran que, para esos casos, existen plegados que conducen a mejores diseños combinados que el plegado de un solo factor o de todos los factores a la vez ya que separan más interacciones dobles confundidas entre sí.

En el año 2002, Li y Mee publicaron un artículo que esencialmente muestra que para muchos diseños de resolución *III* existe más de un plegado que incrementa la resolución a *IV*. Los autores presentan un algoritmo para determinar si un diseño  $2^{k-p}$  de resolución *III* dado, tiene un plegado alternativo al tradicional que no sólo conduce a un diseño combinado de resolución *IV*, sino que separa más interacciones dobles que el plegado tradicional.

Por otra parte, Li y Lin (2003) desarrollaron un método general que permite encontrar los mejores plegados desde el punto de vista del criterio de aberración para cualquier diseño  $2^{k-p}$ , es decir los diseños plegados que conducen a un diseño combinado que tiene mínima aberración entre todos los diseños  $2^{k-p+1}$ . Cabe mencionar que, para un diseño  $2^{k-p}$  dado, existen muchos plegados que conducen al mismo plan experimental, llamados diseños plegados equivalentes.

## APLICACIONES

A continuación, se presentan dos problemas cuya resolución implica el uso de diseños experimentales. Se plantea el uso de fracciones factoriales como diseños iniciales y, dada la necesidad de separar efectos importantes confundidos, se construyen los diseños plegados mediante la estrategia tradicional y el método propuesto por Li y Lin (2003).

### Primer caso

Mee (2009) presenta los resultados de un experimento cuyo objetivo era investigar la capacidad de unión de una sustancia neuropeptídica (SP) modificada por la sustitución de los L-aminoácidos con D-aminoácidos en 9 posiciones de la SP. Los factores A1-A9 (simbolizados de aquí en adelante con las letras de la *A* hasta la *H* y la *J*) indican el aminoácido utilizado en cada posición (+1: L-aminoácido; -1: D-aminoácido). La variable respuesta fue el porcentaje de inhibición del reactivo y el objetivo del experimento fue maximizarla.

La ejecución de un diseño factorial completo  $2^9$  implica la realización de 512 pruebas. Sin embargo, como se mencionó anteriormente, se puede ejecutar una fracción del diseño completo. En este caso, se llevó a cabo, una fracción  $2^{9-4}$  y dada la necesidad de aumentar el número de pruebas, se realizaron dos plegados adicionales:

- $2^{9-4}$  + plegado tradicional: invertir todos los factores
- $2^{9-4}$  + plegado propuesto: invertir dos columnas cualesquiera de los factores generados

Tabla 2. Resultados del diseño  $2^{9-4}$  inicial, plegado tradicional y plegado propuesto

Característica	Inicial	Inicial + plegado tradicional	Inicial + plegado propuesto
Nº de ensayos	32	64	64
Vector W	$W = (0, 8, 7, 0, 0, 0)$	$W = (0, 6, 0, 0, 0, 1)$	$W = (0, 2, 4, 0, 0, 1)$
Resolución	IV	IV	IV
Efectos principales	- 9 efectos confundidos con interacciones triples y de orden superior	- 1 efecto libre de confusión - 8 efectos confundidos con interacciones triples y de orden superior	- 9 efectos confundidos con interacciones triples y de orden superior
Interacciones dobles	- 12 confundidas con otras interacciones dobles e interacciones de orden superior - 8 confundidas con interacciones triples y de orden superior - 4 confundidas entre sí	- 12 confundidas con otras interacciones dobles e interacciones de orden superior - 8 confundidas con interacciones cuádruples - 4 confundidas entre sí	- 6 confundidas con otras interacciones dobles e interacciones de orden superior - 24 confundidas con interacciones triples y de orden superior

## Segundo caso

Dean, Voss y Draguljic (2017) presentan los datos de un estudio llevado a cabo en el año 1989 por S. Lewis, B. Hodgson, R. New y C. Sexton. El objetivo del estudio fue mejorar el rendimiento de un inclinómetro, que es un instrumento que registra el ángulo de inclinación de un objeto como un brazo de grúa. El propósito del experimento fue variar los tamaños relativos de las partes del inclinómetro para encontrar una combinación de factores que reducirían el movimiento. El objetivo del experimento fue seleccionar las combinaciones de los niveles de los factores (longitud de flexión, grosor, tolerancia, etc.) que conducen al menor movimiento, siendo ésta la variable respuesta del estudio.

En este caso, se llevó a cabo, una fracción  $2^{7-3}$  (16 pruebas) y dada la necesidad de aumentar el número de pruebas, se realizaron plegados adicionales:

- $2^{7-3}_{IV}$  + plegado tradicional: invertir todos los factores
- $2^{7-3}_{IV}$  + plegado propuesto: invertir los niveles de cualquiera de los 3 generadores o cualquier combinación de a 2 de los mismos.

Además, el algoritmo de Li y Lin permite mostrar que hay algunas situaciones en las que es preferible considerar diseños iniciales que no sean de máxima resolución pero que conduzcan a un diseño combinado mejor. Por ello, se considera nuevamente el segundo problema presentado en esta tesina y se utiliza como diseño inicial una fracción  $2^{7-3}$ , dando lugar a dos plegados adicionales:

- $2^{7-3}_{III}$  + plegado tradicional: invertir todos los factores

- $2_{III}^{7-3}$  + plegado propuesto: invertir los niveles de los 3 generadores simultáneamente.

Tabla 3. Resultados del diseño  $2_{IV}^{7-3}$  inicial, plegado tradicional y plegado propuesto

Característica	Inicial (IV)	Inicial + plegado tradicional	Inicial + plegado óptimo
Nº de ensayos	16	32	32
Vector W	$W = (0,7,0,0,0)$	$W = (0,7,0,0,0)$	$W = (0,3,0,0,0)$
Resolución	IV	IV	IV
Efectos principales	- 7 efectos confundidos con interacciones triples y de orden superior	- 7 efectos confundidos con interacciones triples y de orden superior	- 1 efecto libre de confusión - 6 efectos confundidos con interacciones triples y de orden superior
Interacciones dobles	- 7 grupos de a 3 confundidas entre sí	- 7 grupos de a 3 confundidas entre sí	- 6 confundidas con otras interacciones dobles e interacciones de orden superior - 6 confundidas con interacciones cuádruples - 3 confundidas entre sí

Tabla 4. Resultados del diseño  $2_{III}^{7-3}$  inicial, plegado tradicional y plegado propuesto

Característica	Inicial (III)	Inicial + plegado tradicional	Inicial + plegado óptimo
Nº de ensayos	16	32	32
Vector W	$W = (2,3,2,0,0)$	$W = (0,3,0,0,0)$	$W = (0,1,2,0,0)$
Resolución	III	IV	IV
Efectos principales	- 5 efectos confundidos con interacciones dobles y de orden superior - 2 efectos confundidos con interacciones triples y superior	- 5 efectos confundidos con interacciones triples y superior - 1 efecto confundido con interacciones superiores a orden 3	- 7 efectos confundidos con interacciones triples o de orden superior
Interacciones dobles	- 6 confundidas con efectos principales - 5 confundidas con interacciones otras interacciones dobles y de orden superior - 2 confundidas con interacciones triples y de orden superior - 3 confundidas entre sí	- 6 confundidas con interacciones otras interacciones dobles y de orden superior - 6 confundidas con interacciones cuádruples y de orden superior - 3 confundidas entre sí	- 1 confundida con otra interacción doble y de orden superior - 2 confundidas con otras interacciones dobles entre sí - 1 confundida con otra interacción doble y triple - 15 confundidas con interacciones triples y de orden superior

## CONCLUSIONES

La conclusión general de este trabajo es que para la mayoría de los diseños existen mejores diseños plegados que el consistente en invertir los signos de todos los factores. Para los diseños de resolución III, el plegado tradicional produce un diseño combinado de resolución IV. El método de Li y Lin mejora el diseño combinado en dos aspectos. Primero, puede incrementar la resolución. Segundo, puede conducir a un diseño de resolución IV con menos palabras de longitud 4, es decir, a menos pares de interacciones dobles confundidas entre sí.

En el caso de diseños de resolución IV, el objetivo principal es separar estimaciones entre pares de interacciones dobles. El método de Li y Lin conduce a diseños combinados que mejoran sustancialmente la estrategia tradicional con respecto a esta cuestión. Estas fortalezas serán aprovechadas siempre que se asignen las columnas de la matriz del diseño correspondiente a aquellos efectos potencialmente significativos, de modo que aparezcan confundidos con interacciones triples o de orden superior.

Los diseños iniciales, en general, son elegidos de acuerdo con el criterio de mínima aberración. Un hallazgo importante en esta investigación es que, en muchas situaciones, comenzar con un diseño con tal característica no necesariamente conduce al mejor diseño combinado. Una mejor opción consiste en elegir el diseño inicial de modo que el combinado sea el de mínima aberración.

Cabe señalar que los paquetes computacionales comúnmente utilizados en el análisis de experimentos en el contexto industrial como Minitab y JMP sólo ofrecen las alternativas tradicionales para la construcción de diseños plegados. Por lo tanto, el uso de los métodos alternativos requiere la implementación de los algoritmos desarrollados.

Una desventaja de los diseños plegados presentados en este trabajo es que el tamaño del diseño puede resultar demasiado grande para algunas situaciones. En estos casos, los diseños plegados parciales desarrollados recientemente pueden resultar una mejor opción. Por otra parte, el criterio de mínima aberración no es el único existente en la bibliografía para evaluar y comparar diseños. En los casos en los que su uso esté justificado, el método de Li y Lin es una opción válida. Sin embargo, en los casos en los que este criterio no sea el adecuado, sería recomendable considerar los mejores diseños plegados con respecto a otros criterios. Estas cuestiones resultan líneas de investigación futuras.

## BIBLIOGRAFÍA

- Dean, Angela, Voss, Daniel, Draguljic, Danel (2017) "Design and Analysis of Experiments". Springer.
- Lawson, John (2015). "Design and Analysis of Experiments with R". CRC Press.
- Li, Huo y Mee, Robert W. (2002) "Better Foldovers Fractions for Resolution III  $2^{k-p}$  Designs". Technometrics.
- Li, William y Lin Dennis K. J. (2003). "Optimal foldover Plans for Two-Level Fractional Factorial Designs". Technometrics.
- Mee, Robert W. (2009) "A Comprehensive Guide to Factorial Two-Level Experimentation". Springer.
- Montgomery, Douglas C. y Runger, George C. (1996). "Foldover of  $2^{k-p}$  Resolution IV Experimental Designs". Journal of Quality Technology.

Oehlert, Gary W. (2010). "A First Course in Design and Analysis of Experiments". Gary W. Oehlert.

Wu, C. F. Jeff y Hamada Michael S. (2009). "Experiments. Planning, Analysis and Optimization". John Wiley&Sons, Inc.

# ANÁLISIS DEL TIEMPO HASTA LA CURACIÓN DE ÚLCERAS VENOSAS MEDIANTE MODELOS LINEALES GENERALIZADOS PARA DATOS CENSURADOS

LIC. JULIA MILLET

Directora: LIC. GUILLERMINA HARVEY

---

Un problema de salud grave, principalmente en la población adulta, es la úlcera vascular. La gran mayoría se produce como consecuencia de la insuficiencia venosa crónica y se denominan úlceras venosas. El tratamiento de los pacientes con esta condición consiste en el uso de vendajes de compresión, distinguiéndose cuatro diferentes terapias: elástica, inelástica, triple venda y bota UNNA. Resulta de interés comparar la eficiencia de estos tratamientos, evaluando cuál logra el menor tiempo promedio hasta la curación.

Dado que el tiempo hasta la curación está registrado como número de meses, se recurre al ajuste de Modelos Lineales Generalizados (MLG), considerando distribuciones Poisson y Binomial Negativa para la variable respuesta. Al contar con datos censurados, es decir, pacientes para los cuales no pudo registrarse el tiempo exacto hasta la curación debido a su fallecimiento, es necesaria una modificación del método de estimación de los MLG de manera que se aplica una adaptación de la función de verosimilitud correspondiente a los MLG de conteo.

El modelo elegido fue el Binomial Negativo bajo el cual se determinó un efecto significativo del tamaño de la úlcera en relación al tiempo hasta la curación, no pudiendo encontrarse diferencias significativas entre los diversos tratamientos.



## INTRODUCCIÓN

Un grave problema de salud, que afecta principalmente a personas mayores, es la úlcera vascular. La lesión aparece cuando la piel se rompe y se localiza encima del tobillo. Esta afección no sólo ocasiona problemas en la salud de los individuos afectados, sino también en su calidad de vida, ya que por tratarse de una lesión de carácter crónico y recidivante los cuidados y tratamientos del paciente son prolongados. Aproximadamente el 85% de las úlceras vasculares se producen debido a insuficiencia venosa crónica, una enfermedad relacionada directamente con la hipertensión venosa y se denominan úlceras venosas (Gómez Ayala, 2008). El tratamiento de los pacientes con esta condición se basa en el uso de vendajes de compresión, los cuales ejercen presión sobre la extremidad ayudando a disminuir la hipertensión venosa. La terapia de compresión puede ser tanto elástica como inelástica (se diferencian en la capacidad que tiene la venda para estirarse) o resultar en una combinación de ambas, lo que se conoce como triple venda. Además, existe otra terapia conocida como bota UNNA que consiste en usar un vendaje común sobre el cual se aplica una solución, de aspecto similar a una gelatina, que a temperatura ambiente se solidifica (Azcona, 2008).

Dado que la úlcera venosa provoca un deterioro en la calidad de vida de quienes la padecen, es de gran importancia que el tratamiento aplicado logre la cura en el menor tiempo posible. A fin de comparar la eficiencia de estos cuatro tratamientos se realizó un estudio en un grupo de pacientes con úlcera venosa y se evaluó el tiempo que demoraron en curarse bajo las diferentes terapias de compresión, con el objetivo de investigar cuál tratamiento logra el menor tiempo promedio hasta la curación.

## LOS DATOS

Se cuenta con la información de 78 pacientes con úlcera venosa tratados en una institución privada de la ciudad de Rosario entre 2015 y 2017. Los mismos fueron controlados desde la aplicación del tratamiento hasta la curación de la úlcera.

Para cada paciente se registró el tamaño de la úlcera (Pequeña - Regular - Mediana - Grande), el tratamiento administrado, la fecha de inicio del tratamiento y la fecha de curación. Dado que los pacientes se realizaban un control en forma mensual, la duración del tratamiento se registró como número de meses transcurridos desde el comienzo del mismo hasta la curación. Cabe destacar que un paciente falleció por causas ajenas al estudio y, por lo tanto, no se conoce el tiempo exacto de curación.

## METODOLOGÍA

Por la forma en la que se diseñó el estudio, el tiempo se registró como una variable de conteo, por lo que es natural recurrir al ajuste de Modelos Lineales Generalizados (MLG), que permiten considerar variables respuesta con distribuciones Poisson y Binomial Negativa, entre otras. Por otra parte, el conjunto de datos disponible presenta observaciones censuradas, es decir, no se conoce para todos los pacientes cuál es el valor exacto del tiempo hasta la curación. En general, el tiempo hasta que se presenta un evento es continuo y se recurre al análisis de supervivencia para tratar este tipo de datos. Sin embargo, en situaciones en las que el tiempo se presenta de

manera discreta, no puede aplicarse el tradicional análisis de datos de supervivencia. Un enfoque posible consiste en el ajuste de MLG para datos de conteo haciendo una adaptación al método de estimación para considerar la presencia de datos censurados.

### Descripción de los Modelos Lineales Generalizados

Un MLG se define en términos de un conjunto de variables aleatorias independientes  $Y_i$ , con  $i = 1, \dots, n$ , cada una con una distribución perteneciente a la familia exponencial que depende de un único parámetro y un conjunto de  $(p - 1)$  variables explicativas. Los parámetros del modelo se relacionan con las variables explicativas mediante el predictor lineal y éste se relaciona con las medias  $\mu_i$ , suponiendo que  $E(y_i) = \mu_i$  a través de alguna función de  $\mu_i$  llamada función de enlace.

Para obtener las estimaciones de los parámetros del modelo se recurre al método de Máxima Verosimilitud (MV). El mismo consiste en maximizar la función de log-verosimilitud,  $l$ , la cual resulta igual a la sumatoria de los logaritmos de las funciones de densidad de probabilidad de las  $y_i$  observadas (Hilbe, 2014).

Por otra parte, para la evaluación del modelo puede utilizarse la estadística *Deviance*,  $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$  (compara la función de log verosimilitud del modelo de interés con la del modelo saturado) o la estadística Chi-Cuadrado de Pearson,  $X^2$ . Ambas estadísticas se distribuyen asintóticamente como una Chi-Cuadrado y sus grados de libertad (gl) son iguales a la diferencia entre el número de parámetros del modelo saturado y el modelo de interés (Agresti, 2015).

### Modelos Lineales Generalizados para datos de conteo

Cuando se analiza una variable respuesta discreta cuyos resultados son conteos, tal como se mencionó, es común utilizar un MLG que asuma la distribución Poisson para el componente aleatorio. Bajo dicha distribución la variable aleatoria  $Y$  se define como el número de veces que ocurre un evento y  $\mu$  es su valor esperado. Sean  $y_1, y_2, \dots, y_n$  observaciones independientes con  $E(y_i) = \mu_i$  y  $x_1, x_2, \dots, x_p$  un conjunto de variables explicativas, el MLG que utiliza el enlace logaritmo para relacionar las medias  $\mu_i$  con el predictor lineal resulta:

$$\log(\mu_i) = \sum_{j=1}^p \beta_j x_{ij} = \beta_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad i = 1, \dots, n.$$

En la práctica, suele ocurrir que la variabilidad observada en los datos excede la supuesta por el modelo, fenómeno que se conoce como sobredispersión. Particularmente en datos ajustados por modelos Poisson, donde la variancia es igual a la media, es frecuente la sobredispersión (Agresti, 2015).

La existencia de sobredispersión puede evaluarse a través del cociente entre la estadística  $X^2$  sobre sus gl, donde valores del cociente mayores a 1,25 indicarán que el modelo presenta este problema. También puede evaluarse a través del test del Multiplicador de Lagrange cuya estadística, bajo la hipótesis nula de no sobredispersión, tiene una distribución Chi-Cuadrado con 1 gl.

Muchas veces la sobredispersión es “aparente” y puede desaparecer si se hacen ciertas modificaciones en el modelo. Sin embargo, si luego de las modificaciones no se

corrige existen distintos métodos para tratarla. El método más habitual consiste en utilizar un MLG Binomial Negativo, ya que esta distribución cuenta con un parámetro adicional, parámetro de dispersión, que permite considerar el exceso de variabilidad presente en los datos (Hilbe, 2014).

La distribución de probabilidad Binomial Negativa puede considerarse como un modelo mezcla en el que se supone que la variable aleatoria  $Y$  tiene distribución Poisson con parámetro  $\lambda$  y que  $\lambda$  varía aleatoriamente de acuerdo a una distribución Gamma. La mezcla de ambas distribuciones da como resultado una distribución Binomial Negativa para  $Y$  con  $E(y) = \mu$  y  $var(y) = \mu + \gamma\mu^2$ , siendo  $\gamma > 0$  el parámetro de dispersión (Agresti, 2015).

Asumiendo el parámetro de dispersión  $\gamma$  constante para las  $n$  observaciones, la expresión del modelo que considera como función de enlace el logaritmo de la media resulta idéntica a la expresión del MLG Poisson descrito previamente.

### Modelos Lineales Generalizados para datos de conteo con censura

La problemática analizada requiere considerar modelos que incluyan observaciones censuradas, es decir, observaciones para las cuales no es posible conocer el tiempo exacto de ocurrencia del evento de interés. Si bien existen tres tipos de censura, en virtud de las características del problema en cuestión sólo puede presentarse uno de ellos: la censura por derecha, la cual ocurre cuando al finalizar el estudio algunos individuos no han presentado el evento o cuando se pierde el seguimiento del paciente, por lo que el tiempo real hasta la ocurrencia del evento será mayor que el observado. Se requiere entonces adaptar la función de verosimilitud correspondiente a los MLG de conteo de manera de tener en cuenta la presencia de observaciones censuradas.

Para el ajuste de un MLG, la función de verosimilitud,  $L$ , dado un conjunto de  $n$  observaciones resulta igual a la productoria de las funciones de densidad de probabilidad de las  $y_i$  observadas. Si se considera una situación en la cual se supone que  $r$  de los  $n$  individuos presentan el evento en los tiempos  $y_1, y_2, \dots, y_r$ , mientras que los tiempos de los  $n - r$  individuos restantes  $y_1^*, y_2^*, \dots, y_{n-r}^*$  se encuentran censurados por derecha se puede reescribir  $L$  para considerar en forma separada los eventos de las observaciones censuradas. Los  $r$  tiempos para los cuales se presentó el evento contribuyen a la función de verosimilitud a través de la función de densidad de probabilidad,  $f(y)$ , mientras que la información de las  $n - r$  observaciones censuradas contribuye a la función de verosimilitud a través de  $1 - F(y^*)$ , donde  $F(y)$  es la función de distribución acumulada, es decir, si una observación es censurada en el tiempo  $y^*$  se conoce que el tiempo hasta la ocurrencia del evento para ese individuo es mayor que  $y^*$  con probabilidad  $P(Y > y^*) = 1 - F(y^*)$ . Por lo tanto, la función de verosimilitud de las  $n$  observaciones resulta:

$$L = \prod_{j=1}^r f(y_j) \prod_{k=1}^{n-r} [1 - F(y_k^*)].$$

El primer factor considera los  $r$  tiempos en los que se presentó el evento y el segundo factor considera los  $n - r$  tiempos censurados. Con el fin de reducir la expresión anterior se define la variable indicadora  $\delta_i$  que toma el valor 0 cuando el tiempo  $y_i$  es

censurado y toma el valor 1 cuando no lo es, de manera que la función de verosimilitud puede expresarse (Collett, 2003):

$$L = \prod_{i=1}^n \{f(y_i)\}^{\delta_i} \{1 - F(y_i)\}^{1-\delta_i}.$$

De este modo, para contemplar las observaciones censuradas se debe hacer una modificación en las funciones de verosimilitud utilizadas habitualmente, reemplazando en la expresión anterior la  $f(y)$  y la  $F(y)$  de acuerdo a la forma correspondiente a la distribución Poisson o Binomial Negativa, según sea el caso del modelo que se ajuste. Para obtener las estimaciones MV de los parámetros desconocidos se utiliza el logaritmo de  $L$ .

### Selección de modelos

Cuando el interés es comparar un modelo con otro más complejo que incluye parámetros adicionales, puede realizarse un test de razón de verosimilitud (RV) que evalúe la hipótesis de que el modelo más simple ajusta, utilizando como estadística del test la diferencia entre la *Deviance* de cada modelo.

Por otra parte, en cuanto a la selección del MLG Poisson o Binomial Negativo, puede realizarse un test de hipótesis que evalúe si el parámetro de dispersión de la distribución Binomial Negativa,  $\gamma$ , es significativamente distinto de cero, ya que la distribución Poisson es el caso límite de la distribución Binomial Negativa cuando  $\gamma \rightarrow 0$  (Agresti, 2015).

## RESULTADOS

### Descripción de los datos

Los 78 pacientes estudiados fueron tratados, como ya se dijo, con diferentes terapias de compresión. La mediana del tiempo hasta la curación de la úlcera venosa resultó igual a 2 meses en todos los casos. Sin embargo, aquellos sometidos al tratamiento con bota UNNA presentan una mayor variabilidad (Tabla 1).

Tabla 1. Número de pacientes, mediana y rango intercuartil del tiempo hasta la curación de la úlcera venosa (en meses) según tratamiento (n=78)

Tratamiento	Número de pacientes (%)	Mediana	Rango Intercuartil
Venda elástica	13 (17%)	2	1 – 3,0
Venda inelástica	22 (28%)	2	1 – 3,0
Triple venda	32 (41%)	2	1 – 3,0
Bota UNNA	11 (14%)	2	2 – 6,5

Con respecto al tamaño de la úlcera, tal como es de esperar los pacientes que tienen úlceras grandes son los que presentan un tiempo mediano hasta la curación mayor. Para estos pacientes el tiempo mediano es de 5 meses, mientras que para el resto de los pacientes los tiempos medianos hasta la curación son iguales a 2 meses

aproximadamente. Además, los pacientes con úlceras grandes muestran una variabilidad superior a la del resto (Tabla 2).

Tabla 2. Número de pacientes, mediana y rango intercuartil del tiempo hasta la curación de la úlcera venosa (en meses) según tamaño de la misma (n=78)

Tamaño de la úlcera venosa	Número de pacientes (%)	Mediana	Rango Intercuartil
Pequeña (Superficie $\leq 3$ cm <sup>2</sup> )	29 (37%)	2	1,0 – 2,0
Regular (3 cm <sup>2</sup> < Superficie $\leq 4$ cm <sup>2</sup> )	16 (21%)	2	1,0 – 2,0
Mediana (4 cm <sup>2</sup> < Superficie $\leq 6$ cm <sup>2</sup> )	18 (23%)	2	1,3 – 3,0
Grande (Superficie > 6 cm <sup>2</sup> )	15 (19%)	5	2,5 – 8,5

Cabe recordar que hay una observación censurada que corresponde a un paciente que falleció por causas ajenas al estudio y que durante el transcurso de 5 meses no había alcanzado la curación. El mismo presentaba una úlcera grande y había sido asignado al tratamiento triple venda.

### Ajuste de modelos

A continuación, se presentan los distintos ajustes realizados para modelar el tiempo medio hasta la curación de la úlcera venosa.

### Ajuste del modelo Poisson

Inicialmente se ajusta un modelo Poisson con las variables explicativas disponibles, tratamiento y tamaño de la úlcera. Los resultados obtenidos para el ajuste del modelo se presentan en la Tabla 3.

Tabla 3. Parámetros estimados, errores estándares y probabilidades asociadas al test de Wald para el modelo Poisson

Parámetro	Estimación	Error estándar	Probabilidad asociada
Intercepto	0,7359	0,1898	0,0002
Tratamiento Inelástica	-0,3620	0,2215	0,1066
Tratamiento Triple Venda	-0,1299	0,1973	0,5125
Tratamiento UNNA	0,1818	0,2166	0,4042
Tamaño Regular	-0,0938	0,2466	0,7049
Tamaño Mediano	0,2198	0,2078	0,2936
Tamaño Grande	1,2579	0,1797	<0,0001

Se realizan los test de RV que evalúan globalmente el efecto de cada covariable, concluyéndose que el efecto del tamaño de la úlcera resulta significativo ( $p < 0,0001$ ), mientras que el efecto del tratamiento no ( $p = 0,2713$ ). Se evalúa la presencia de sobredispersión en el modelo mediante el cociente entre la estadística  $X^2$  sobre sus gl y mediante el test del Multiplicador de Lagrange; se concluye que este fenómeno está presente en los datos ( $p < 0,0001$ ).

Debido a que puede suceder que la sobredispersión sea aparente y se corrija al agregar algún término faltante al predictor del modelo, se considera la inclusión del término de interacción entre tratamiento y tamaño de la úlcera. El test de RV correspondiente al efecto de la interacción no resulta significativo ( $p=0,8963$ ), por lo que se decide ajustar un modelo Binomial Negativo considerando sólo los efectos principales.

### Ajuste del modelo Binomial Negativo

Los resultados obtenidos en el ajuste del modelo se presentan en la Tabla 4.

Se realizan nuevamente los test de RV para evaluar globalmente el efecto de cada covariable y al igual que lo observado en el ajuste anterior, resulta significativo sólo el efecto del tamaño de la úlcera ( $p=0,0002$ ) y no así el del tratamiento ( $p=0,5828$ ). Por otra parte, la estadística asociada al test de RV para el parámetro de dispersión de la distribución Binomial Negativa arroja una probabilidad asociada de 0,0030, es decir, el parámetro adicional de la distribución Binomial Negativa es significativo. Por lo tanto, el MLG para datos censurados con distribución Binomial Negativa es el elegido para modelar los datos bajo análisis.

Tabla 4. Parámetros estimados, errores estándares y probabilidades asociadas al test de Wald para el modelo Binomial Negativo

Parámetro	Estimación	Error estándar	Probabilidad asociada
Intercepto	0,7110	0,2164	0,0016
Tratamiento Inelástica	-0,3032	0,2556	0,2396
Tratamiento Triple venda	-0,0871	0,2292	0,7051
Tratamiento UNNA	0,1803	0,2630	0,4952
Tamaño Regular	-0,1039	0,2674	0,6987
Tamaño Mediano	0,2119	0,2280	0,3558
Tamaño Grande	1,2565	0,2102	<0,0001

### Interpretación del Modelo Binomial Negativo

En función del ajuste del modelo Binomial Negativo, pueden estimarse los tiempos medios hasta la curación de la úlcera venosa según el tratamiento asignado al paciente y el tamaño de la úlcera. Estos valores se presentan en la Tabla 5.

Los mayores tiempos medios estimados corresponden a aquellos pacientes con úlceras grandes. Más específicamente, mientras que en estos pacientes los tiempos medios estimados de curación fluctúan entre 5 y 8 meses, las úlceras pequeñas, regulares y medianas demorarán en curarse entre 1 y 3 meses.

Tabla 5. Tiempos medios, en meses, estimados hasta la curación de la úlcera venosa según tratamiento y tamaño de la úlcera

Tratamiento	Tamaño de la úlcera			
	Pequeña	Regular	Mediana	Grande
Venda elástica	2,04	1,84	2,51	7,10
Venda inelástica	1,51	1,36	1,86	5,26
Triple venda	1,88	1,70	2,32	6,55
Bota UNNA	2,44	2,20	3,00	8,50

Para tener una medida del efecto que tiene el tamaño de la úlcera sobre la demora en la curación pueden calcularse las razones de tiempos medios estimados. Por ejemplo, para pacientes que recibieron un mismo tratamiento, el tiempo medio estimado hasta la curación de la úlcera venosa cuando la misma es grande es 3,5 veces el tiempo hasta la curación de los pacientes que presentan úlceras pequeñas:

$$\frac{\hat{\mu}_{UG}}{\hat{\mu}_{UP}} = \exp(1,25) = 3,50 .$$

## CONSIDERACIONES FINALES

En este trabajo se analizaron datos correspondientes a 78 pacientes con úlcera venosa en la pierna, atendidos en una institución privada de la ciudad de Rosario entre los años 2015 y 2017. Los pacientes fueron tratados con distintas terapias de compresión y se controlaron en forma mensual desde la aplicación del tratamiento hasta la curación de la úlcera. Un paciente falleció por causas ajenas al estudio y, por lo tanto, no pudo registrarse el tiempo hasta su curación.

La presencia de observaciones censuradas suele ser habitual cuando se sigue a un grupo de pacientes durante un período de tiempo. Es importante incorporar la información que se tiene de estos pacientes, ya que de no hacerlo es probable arribar a estimaciones sesgadas, por ello en este trabajo se recurrió al ajuste de MLG para datos de conteo censurados, de modo de poder contemplar la información aportada por el paciente que falleció. Si bien en esta oportunidad sólo se contaba con una observación censurada y su exclusión para aplicar un MLG clásico no trae diferencias sustanciales en los resultados, se considera importante tener en cuenta esta adaptación de la metodología usual de ajuste de un MLG que incorpora datos censurados.

El modelo elegido fue el Binomial Negativo, bajo el cual no pudieron encontrarse diferencias significativas entre los diversos tratamientos posiblemente debido a la cantidad de pacientes en los grupos evaluados.

Con el objeto de enriquecer los resultados, sería importante registrar para futuros estudios más información de cada paciente como la edad, el sexo, si la úlcera se trataba de un caso de recidiva o no, entre otras características.

Por último, cabe mencionar que cuando se cuenta con variables respuesta que miden el tiempo hasta la ocurrencia de un evento de interés es natural recurrir al análisis de datos de supervivencia, ya que usualmente, el tiempo está registrado en forma continua. Sin embargo, en ciertas ocasiones, se dispone de los tiempos hasta el evento

de interés medido en forma discreta. Algunos enfoques alternativos para estos casos son los modelos de *odds* proporcionales o los modelos basados en funciones *hazard* para tiempos discretos. Sería interesante continuar con el análisis de las posibles diferencias en la aplicación de estos diversos enfoques.

## BIBLIOGRAFÍA

- Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Azcona, L. (2008). Insuficiencia venosa: Prevención y tratamiento. Farmacia Profesional, 22(10), 7-59. Recuperado de <http://www.elsevier.es/es-revista-farmacia-profesional-3-articulo-insuficiencia-venosa-prevencion-tratamiento-13129193>.
- Collett, D. (2003). Modelling Survival Data in Medical Research (Second Edition). Londres, Inglaterra: Chapman & Hall/CRC.
- Gómez Ayala, A. E. (2008). Úlceras vasculares: Factores de riesgo, clínica y prevención. Farmacia Profesional, 22(6), 7-65. Recuperado de <http://www.elsevier.es/es-revista-farmacia-profesional-3-articulo-ulceras-vasculares-factores-riesgo-clinica-13124067>
- Hilbe, J.M. (2014). Modeling Count Data. Cambridge, Inglaterra: Cambridge University Press.

# CONSTRUCCIÓN Y ANÁLISIS DE UN DISEÑO $\alpha$ PARA LA COMPARACIÓN DE CULTIVARES DE TRIGO

LIC. CAMILA MIOTTI

Directora: LIC. MARÍA BELÉN CONDE

Co-Directora: MG. CRISTINA CUESTA

---

En INTA (Instituto Nacional de Tecnología Agropecuaria) se lleva a cabo un programa de mejoramiento de trigo el cual tiene como objetivo seleccionar, evaluar y comparar un conjunto de genotipos para inscribirlos como nuevas variedades en el mercado.

Para dicho fin se realizan ensayos regionales y el diseño utilizado en cada localidad es muy importante para obtener resultados más eficientes. Dado que en cada localidad se presentan características geográficas diferentes y que el diseño a utilizar debe contemplarlas a todas, se utiliza una clase de diseños en bloques incompletos resolubles conocidos como Diseños  $\alpha$  que se adaptan a los requerimientos locales.

El análisis estadístico de dichos diseños se realiza mediante modelos mixtos y modelos con efectos fijos con correlación espacial. Ambos enfoques tienen como objetivo tener en cuenta la falta de independencia entre las observaciones, producida naturalmente por la variabilidad del suelo.

La finalidad de este análisis es obtener un orden de los genotipos en cuanto a su rendimiento y encontrar los de mayor rinde a fin de que los investigadores puedan luego compararlos con los resultados de otras regiones.



## INTRODUCCIÓN

En el Instituto Nacional de Tecnología Agropecuaria (INTA) existe un *programa de mejoramiento de trigo*, el cual tiene como objetivo determinar la adaptabilidad, sanidad y calidad del germoplasma. En dicho programa se pasa por diferentes etapas hasta lograr seleccionar un conjunto de genotipos, los cuales necesitan ser comparados con un grupo de testigos de comportamiento conocido, es decir variedades que ya están incorporadas en el mercado, y esta comparación resulta trascendente para decidir sobre la posible inscripción de una nueva variedad al mercado.

Para llevar a cabo la selección y comparación de los genotipos, INTA utiliza 8 localidades de evaluación: Roque Sáenz Peña, Paraná, Marcos Juárez, Corral de Bustos, Pergamino, Balcarce, Barrow y Bordenave en Buenos Aires; y a este tipo de ensayos se los denomina Regionales. Del análisis conjunto de la información proveniente de estos ensayos se decide anualmente la presentación de nuevas variedades.

La metodología estadística es una de las herramientas fundamentales para llevar a cabo el programa ya que los diseños de experimentos y un análisis adecuado de la información obtenida a partir de los ensayos es indispensable para que el programa sea eficaz.

La elección del diseño a realizar es muy importante, ya que la variación natural que presenta el suelo con respecto a la fertilidad puede conducir a estimaciones y resultados deficientes si no se utiliza el diseño apropiado. Es por ello que se utiliza un Diseño  $\alpha$ , el cual se adapta a los requerimientos locales.

El análisis estadístico de los datos obtenidos a partir de estos diseños se puede llevar a cabo mediante el planteo de distintos modelos. Frente a la diversidad de terreno con el que se cuenta, en general se consideran modelos mixtos o modelos con efectos fijos y correlación espacial.

En este trabajo se estudian y analizan datos sobre los rendimientos obtenidos de 54 genotipos de trigo en la localidad de Bordenave en el año 2015, provenientes de un Diseño  $\alpha$ .

Luego de comparar y seleccionar el mejor modelo que ajusta a los datos, el mismo se utiliza para poder llevar a cabo una comparación entre los distintos genotipos, a fin de obtener un ordenamiento e indicar aquellos que se espera que presenten mayor rendimiento.

## METODOLOGÍA

### Diseños $\alpha$

A menudo, existen ensayos en los cuales el número de tratamientos es fijo y lo suficientemente grande para requerir el uso de diseños en bloques incompletos. Del mismo modo, pueden presentarse otras restricciones, entre ellas que el número de réplicas esté predeterminado y que el diseño sea resoluble (ya que el uso de diseños resolubles permite, en ensayos a campo agrícolas, que la siembra o la cosecha se realicen en etapas).

Los Diseños  $\alpha$  son un tipo de diseño en bloques incompletos resolubles que no tienen limitantes en cuanto al tamaño del bloque excepto por la restricción que el número de tratamientos,  $t$ , debe ser un múltiplo del tamaño del bloque  $k$ , de manera que  $t = sk$ , para tener un diseño resoluble con tamaños de bloques iguales (Kuehl, 2001).

Una ventaja que presentan estos diseños es que son fáciles de construir, los mismos pueden ser generados por sustitución cíclica a partir de un arreglo inicial de números. Por otro lado, los mismos pueden ser generados por diferentes programas computacionales, como por ejemplo Alphagen, CycDesign, Genstat y R, aunque este último presenta algunas limitaciones y sólo permite construir una clase especial de los diseños de interés. Hay otros programas que, si bien no tienen incorporadas las rutinas para construirlos, se puede armar un script para que realice la tarea, tal es el caso de SAS.

### **Metodologías de análisis para datos provenientes de un diseño $\alpha$**

Respecto al análisis de datos provenientes de Diseños  $\alpha$ , se proponen dos posibles metodologías estadísticas enmarcadas dentro de la teoría de modelos lineales: modelos lineales mixtos y modelos de efectos fijos con correlación espacial. En ambos casos se intenta modelar la correlación generada por datos extraídos de parcelas ubicadas cercanas en el espacio. Tener en cuenta estas correlaciones permite obtener estimaciones más precisas y comparaciones más certeras.

### **Modelos Mixtos**

Hay situaciones en las cuales no es adecuado utilizar los modelos lineales clásicos de análisis de variancia y regresión porque no cumplen los supuestos de independencia, igualdad de variancias e incluso linealidad.

En lugar de ellos, se pueden plantear modelos mixtos cuya característica distintiva es que la respuesta media es modelada como una combinación de características poblacionales en común que comparten las unidades (los que se consideran efectos fijos del modelo), y efectos específicos o propios de la unidad (considerados efectos aleatorios) (Fitzmaurice, Laird, & Ware, 2004). Estos modelos logran captar una buena parte de la correlación espacial.

Para llevar a cabo la comparación entre varios modelos propuestos se utilizan los criterios de información Akaike (AIC) y/o Schwarz (BIC).

La estrategia es utilizar el modelo más completo y comparar distintas estructuras de las matrices de covariancias de los efectos aleatorios y/o errores a través de las estadísticas AIC y/o BIC. Una vez seleccionada la estructura de covariancia que minimiza esas estadísticas, se postulan y comparan modelos más parsimoniosos en la parte media utilizando la estructura de covariancias antes seleccionada (también utilizando los criterios de AIC y BIC o bien a través de un test de razón de verosimilitud).

## Modelos de Correlación Espacial

La variabilidad espacial, tal como lo indica su nombre, se refiere a la variación entre las observaciones en el espacio (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006). En los ensayos realizados a campo, es fácil imaginar cómo la variación natural en la fertilidad, profundidad y humedad del suelo pueden provocar una estructura espacial, no aleatoria, en el desarrollo y comportamiento de los cultivos del ensayo. Es por ello que existen otras alternativas de análisis distintas a los métodos estadísticos clásicos, las cuales permiten, medir y controlar dicha correlación espacial presente en el ensayo experimental, ya que la misma supone la violación de uno de los requisitos fundamentales de los métodos de estadística clásica, la independencia de las observaciones.

Para poder utilizar estos modelos es necesario poder identificar la distancia a la que se encuentran las observaciones, ya que en estos modelos la covariancia se define en función de la distancia entre las observaciones, de allí la importancia de obtener un plano que permita establecer la distancia.

Por otra parte, los modelos de correlación espacial pueden clasificarse en dos grandes grupos: *isotrópicos* y *anisotrópicos*. Los modelos isotrópicos consideran la misma correlación en todas las direcciones, mientras que los anisotrópicos consideran distintas correlaciones para dos direcciones principales. Este último modelo es especialmente útil para modelar la correlación espacial en presencia de pendientes, lomas o vientos predominantes en una determinada dirección.

Los modelos de covariancia isotrópicos más utilizados son el Exponencial y el de Potencia.

Al igual que en los Modelos Mixtos, la estrategia es ajustar un modelo maximal para la parte media del modelo, seleccionar la estructura de covariancia adecuada y por último identificar los efectos significativos del modelo.

## Programa de Mejoramiento de Trigo

Una de las actividades que se llevan a cabo en la Estación Experimental del INTA de Marcos Juárez en el marco del Programa de Mejoramiento de Trigo, es evaluar un conjunto de genotipos con el fin de inscribirlos como una nueva variedad comercial en el INASE (Instituto Nacional de Semillas) (Franchina, B, & J, 2004).

La información básica para la realización de este trabajo corresponde a un ensayo a campo de la localidad de Bordenave durante el año 2015 en el cual se obtuvo el rendimiento de 54 genotipos de trigo.

Para poder evaluar y estudiar el rendimiento de los genotipos de interés, los investigadores debieron ubicar en el terreno cada uno de ellos siguiendo el plano de un Diseño  $\alpha$  obtenido a partir del *Software* Alphagen.

Para evaluar si la elección del diseño experimental utilizado fue correcta se realiza una estimación de la Eficiencia Relativa (ER). Para ello se comparan los CME del diseño utilizado con el de un diseño en bloques completos aleatorizados (DBCA). El valor de la ER resulta 2,48, lo que indica que el CME del análisis llevado a cabo considerando un DBCA, es aproximadamente 2,5 veces mayor que el CME

obtenido considerando un Diseño  $\alpha$ . Claramente, la utilización de un Diseño  $\alpha$  brinda mejores resultados y permite realizar estimaciones y comparaciones entre los tratamientos de forma más precisa.

## RESULTADOS

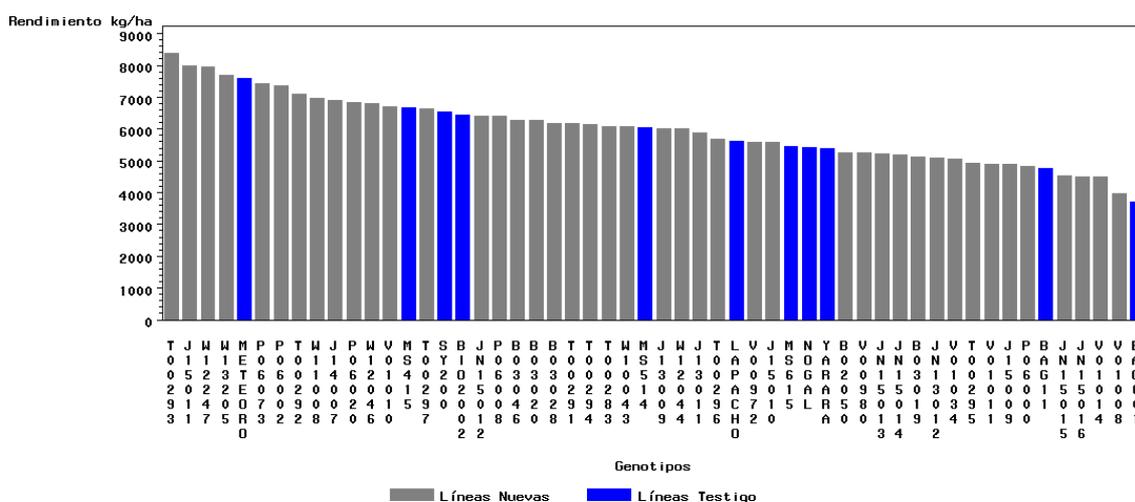
Para comenzar con el análisis de los datos provenientes del ensayo descrito anteriormente se realiza un breve análisis descriptivo, a fin de conocer características básicas de los datos recolectados. Luego se proponen dos caminos alternativos de análisis, uno a través de la utilización de modelos mixtos y otro bajo el enfoque de modelos con correlación espacial.

### Análisis Descriptivo

En la Figura 1 se muestran los promedios de los rendimientos de los genotipos, ordenados de mayor a menor, y además se realiza una discriminación de las nuevas líneas respecto a las variedades testigos, a fin de obtener una interpretación y comparación visual más sencilla. El mayor rendimiento lo presentan cuatro líneas nuevas (T00293, J15011, W12247 y W13205), las cuales parecen superar a una de las variedades comerciales, Meteor. Estas cuatro líneas provienen de las plantas experimentales de Paraná, Marcos Juárez y Barrow y todas ellas presentan sólo un año de antigüedad en el programa, es decir, es el primer año que se ensayan a campo.

Por otro lado, la variedad comercial BAG601 es la que presenta el menor rendimiento y es superada sólo por cuatro líneas nuevas (JN15015, JN15016, V01014 y V01008). Estas cuatro líneas sólo tienen un año de antigüedad en el programa y fueron ensayadas, dos en la localidad de Sáenz Peña y dos en Bordenave, respectivamente.

Figura 1. Rendimiento de los Genotipos de Trigo (kg/ha)



### Análisis a través de Modelos Mixtos

Frente a la diversidad de terreno con el que se cuenta, los efectos repetición, fila y columna se consideran aleatorios; mientras que el efecto genotipo se considera fijo.

Para comenzar con el análisis utilizando modelos mixtos, se consideran distintas estructuras de covariancias para los efectos aleatorios del modelo, logrando así captar de la mejor manera, si es que hubiere, la variabilidad natural del suelo. La elección adecuada de dichas estructuras de covariancias resulta importante ya que, de la misma dependen las estimaciones a realizar. Se consideran tres estructuras de covariancias: Independencia ( $\sigma^2I$ ), Simetría Compuesta (SC) y Autorregresiva de Orden 1 (AR(1)). La segunda de ellas asume que la correlación entre las observaciones dentro de una misma repetición es constante, sin importar la distancia a la que se encuentran, mientras que la tercera considera que la correlación dentro de una repetición decae a medida que la distancia entre las observaciones aumenta (Fitzmaurice, Laird, & Ware, 2004). Para el efecto Repetición sólo se considera la estructura SC y una de Independencia, mientras que para los efectos Fila y Columna se consideran los tres patrones (SC, AR(1) e Independencia). Es decir, se utilizan 2 estructuras para el efecto Repetición, 3 para el efecto Fila y para el efecto Columna, obteniendo así, al combinar todas estas estructuras, un total de 18 modelos.

En la Tabla 1 se presentan los resultados obtenidos al ajustar 9 modelos considerando una estructura de Independencia ( $\sigma^2I$ ) para el efecto Repetición y en la Tabla 2 se presentan los 9 modelos restantes considerando la estructura SC, también para el efecto repetición.

Tabla 1. Resultados de los modelos mixtos ajustados para el Diseño  $\alpha$  considerando una estructura  $\sigma^2I$  para el efecto Repetición

		FILA			
		$\sigma^2I$	SC	AR(1)	
COLUMNA	$\sigma^2I$	AIC:181	AIC:181	AIC:180	
		5,2	7,2	2,7	
		BIC:181	BIC:181	BIC:180	
	SC	2,5	3,6	0	
		AIC:181	AIC:181	No	
		7,2	9,2	Converge	
	AR(1)	BIC:181	BIC:181		
		3,6	4,7		
		AIC:181	AIC:181	AIC:180	
	(1)	7,2	9,2	4,7	
		BIC:181	BIC:181	BIC:180	
		3,6	4,7	1,1	

Tabla 2. Resultados de los modelos mixtos ajustados para el Diseño  $\alpha$  considerando una estructura SC para el efecto Repetición

COLUMNA	FILA			
		$\sigma^2 I$	SC	AR(1)
	$\sigma^2 I$	AIC:1817	AIC:1819	No
		BIC:1813	BIC:1814	Converge
	SC	AIC:1819	AIC:1821	No
		BIC:1814	BIC:1815	Converge
1)	AR(1)	AIC:1819	AIC:1821	No
		BIC:1814	BIC:1815	Converge

Considerando los valores de AIC (y BIC) en las Tablas 1 y 2, se decide que el modelo adecuado para el análisis de este ensayo utilizando modelos mixtos es el que considera una estructura de Independencia, tanto para el efecto repetición como para el efecto columna, y una estructura AR(1) para el efecto fila.

Una vez seleccionadas las estructuras de covariancias se evalúa la significación del efecto fijo considerado en el modelo, resultando estadísticamente significativo ( $p < 0,0001$ ).

Para continuar el análisis a través de modelos mixtos se presentan los rendimientos estimados a través del modelo seleccionado con sus respectivos errores estándares (Tabla 3).

La última columna de la Tabla 3 agrupa las variedades por letras, donde los genotipos que comparten letras no difieren significativamente. Las comparaciones múltiples se realizan a través de la Diferencia Mínima Significativa. Es decir, el genotipo T00293 no difiere significativamente en cuanto al rendimiento del genotipo W12247, pero sí lo hace del genotipo P06073.

Tabla 3. Estimaciones de los rendimientos de los genotipos de trigo

GENOTIPOS	REND. ESTIMADO	ERROR ESTÁNDAR	COMP. MÚLTIPLES
T00293	8038.71	818.22	A
W12247	7774.95	821.51	AB
J15011	7439.31	813.37	ABC
METEORO	7391.61	821.51	ABCD
W13205	7338.78	821.34	ABCDE
P06020	7095.46	816.21	ABCDEF
P06002	6952.00	818.99	ABCDEFG
B03046	6928.81	817.15	ABCDEFGH
P06073	6876.74	815.51	BCDEFGHI
T00292	6774.01	820.22	BCDEFGHIJ

### Análisis a través de Modelos con Correlación Espacial

En esta vía de análisis, se ajustaron 4 modelos de covariancias para los errores: Exponencial Isotrópico, Exponencial Anisotrópico, Potencia Isotrópico y Potencia Anisotrópico.

En este trabajo la distancia entre las observaciones se considera dada por las coordenadas de las filas y las columnas. Si, por ejemplo, se desea obtener la distancia entre la observación 7 y 26 de la repetición 3, tal como puede observarse en la Figura 2, las mismas difieren en una columna y una fila, luego la distancia entre ellas se obtiene calculado la hipotenusa formada por dichas observaciones, siendo esta igual a  $\sqrt{2}$ .

Figura 2. Disposición de las observaciones en el plano de la Repetición 3

Repetición 3 columna	1	2	3	4	5	6	7	8	9	
Fila (Bloque)	1	7	30	48	44	9	46	49	25	5
2	53	26	34	31	47	43	17	41	40	
3	1	12	36	19	35	52	13	28	20	
4	38	50	2	4	27	6	16	42	33	
5	10	18	54	45	23	32	51	29	37	
6	22	14	15	11	8	21	24	3	39	

Los modelos que consideran estructuras Exponencial y Potencia, ambas de la forma Isotrópica, brindan el mismo ajuste (Tabla 4), en el sentido que ambos modelos logran minimizar de igual forma los valores de AIC y BIC. Además, para los datos de este ensayo brinda un mejor ajuste aquel modelo que no tiene en cuenta la correlación espacial que aquel que la considera en forma Anisotrópica.

Tabla 4. Resultados de los modelos ajustados para el Diseño  $\alpha$  considerando estructuras de correlación espacial

<b>ESTRUCTURAS</b>	<b>AIC</b>	<b>BIC</b>
<b>Sin estructura espacial</b>	1412,6	1411,7
<b>Exponencial Isotrópica</b>	1410,6	1408,8
<b>Exponencial Anisotrópica</b>	1415,6	1411,1
<b>Potencia Isotrópica</b>	1410,6	1408,8
<b>Potencia Anisotrópica</b>	1412,2	1409,5

Los resultados se obtuvieron utilizando el programa estadístico SAS 9.4.

Para la interpretación de los resultados resulta más sencillo comprender y describir el modelo con estructura de Potencia que con una Exponencial, es por ello que se continúa el análisis con dicho modelo.

Los test realizados para evaluar la significación de los efectos fijos del modelo muestran que el efecto columna es no significativo ( $p=0,4017$ ). Es por ello que se decide ajustar un nuevo modelo que no considere dicho efecto. Este último modelo brinda un mejor ajuste y todos sus efectos son significativos ( $p < 0,0001$ ).

Junto a las estimaciones se presentan los errores estándares y los resultados obtenidos de las comparaciones múltiples, las cuales se realizaron a través de la Diferencia Mínima Significativa. Nuevamente, los genotipos que comparten letra no difieren significativamente (Tabla 5).

Tabla 5. Estimaciones de los rendimientos de los genotipos de trigo

<b>GENOTIPOS</b>	<b>REND. ESTIMADO</b>	<b>ERROR ESTÁNDAR</b>	<b>COMP. MÚLTIPLES</b>
T00293	7855.36	380.51	A
W12247	7477.90	387.63	AB
J15011	7386.52	386.65	ABC
METEORO	7373.50	379.76	ABCD
W13205	7136.50	379.40	ABCDE
P06073	7059.77	390.53	ABCDEF
P06020	7044.00	385.92	ABCDEFG
B03046	6949.59	390.66	ABCDEFGH
B03028	6917.15	401.56	ABCDEFGHI
T00294	6916.68	407.77	ABCDEFGHI

Los genotipos T00293, W12247, J15011 superan en cuanto al rendimiento al genotipo testigo Meteoro. Además, ambas metodologías de análisis brindan el mismo orden, en cuanto a las estimaciones de los rendimientos, para los primeros 5 genotipos (Tabla 5).

Luego, dado que la utilización de modelos con correlación espacial brinda estimaciones más precisas (los errores estándares son menores) y que los mismos resultan más sencillos para comprender y explicar a los investigadores acerca de cómo actúa la variabilidad espacial sobre las observaciones, se concluye que dichos modelos son más adecuados de implementar que los modelos mixtos.

## DISCUSIÓN

Los datos utilizados en este trabajo corresponden a un ensayo regional de INTA realizado en la localidad de Bordenave en el año 2015, a fin de inscribir nuevas variedades comerciales de trigo. Para evaluar el comportamiento del ensayo realizado en dicha localidad en comparación a las otras regiones experimentales, podría llevarse a cabo un análisis de genotipo-ambiente. A través de este análisis se observarían puntalmente a las líneas T00293, W12247 Y J15011 que en este ensayo tuvieron una excelente performance, evaluando si en otras regiones experimentales se pudo ver reflejado el mismo comportamiento.

Por otro lado, para la realización de este trabajo se utilizan los rendimientos obtenidos de cada uno de los genotipos, pero en los ensayos se miden y registran además otras características y propiedades de cada uno de ellos, como lo son la calidad panadera, la resistencia frente a diversas enfermedades y comportamiento frente a cambios climáticos. Generalmente, dependiendo el interés del investigador, se busca un balance entre el rendimiento, la calidad panadera y la fortaleza de la planta. Es por ello que para lograr un análisis y estudio más completo de los genotipos se sugiere una alternativa de Análisis Multivariado que contemple a todas estas características en conjunto, ya que muchas veces una variedad puede obtener muy buenos resultados en cuanto al rendimiento pero no poseer buena calidad panadera, o bien ser plantas débiles frente a ciertas enfermedades y/o climas no apropiados.

Con respecto a las estimaciones realizadas de los rendimientos de los distintos genotipos de trigo, puede comentarse que las estimaciones obtenidas utilizando Modelos con Correlación Espacial resultan más precisas que las obtenidas utilizando Modelos Mixtos, ya que los errores estándares de las mismas se redujeron, prácticamente al 50%. Además, en cuanto al ordenamiento de los rendimientos, ambas metodologías establecen un mismo orden para los 5 primeros genotipos.

Existen otras alternativas para controlar la correlación espacial, como por ejemplo modificar el nivel de significación a un valor más conservador, ya que reduciendo el mismo se compensa el aumento de probabilidad de cometer un Error Tipo I que genera la autocorrelación espacial. Otra propuesta es utilizar métodos de ajuste por el vecino más próximo, los cuales se basan en corregir los datos según la covariación en las unidades experimentales vecinas, calculando una covariable de ajuste a partir de los residuales en las observaciones vecinas.

Son muchos los estudios que demuestran las ventajas del uso de modelos con errores correlacionados en presencia de correlación espacial. De hecho, este tipo de modelos se empiezan a usar de forma rutinaria en ensayos agronómicos.

## BIBLIOGRAFÍA

- Burbano, T. C., & Escobar, H. C. (2001). Eficiencia relativa del diseño latice con respecto a los diseños bloques al azar e irrestrictamente al azar en un ensayo de rendimiento de maiz de clima medio. *Revista de Ciencias Agrícolas*, 140 - 149.
- Cochran, W., & Cox, G. (1971). *Diseños experimentales*. México: Editorial Trillas.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied Longitudinal Analysis*. Hoboken, New Jersey: John Wiley & Sons, INC.
- Franschina, J., B, M., & J, B. C. (2004, Octubre). Red de Evaluación de Líneas de Trigo Próximas a Inscripción. Bahía Blanca.
- Hinkelmann, K., & Kempthorne, O. (2005). *Design and Analysis of Experiments - Volume 2*. New Jersey: John Wiley & Sons, Inc.
- Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury: Thomson Learning.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for Mixed Models*. Cary, NC, USA: SAS Institute Inc.
- Montgomery, D. C. (2004). *Diseño y análisis de experimentos*. Mexico: Editorial Limusa S.A.
- Patterson, H., & Williams, E. (1976). A New Class of Resolvable Incomplete Block Design. *Biometrika*, 83-92.
- Piepho, H. P., Büchse, A., & Truberg, B. (2006). On the use of multiple lattice designs and a-designs in plant breeding trials. *Plant Breeding* 125, 523-528.
- Zas, R. (2018, Diciembre 12). Autocorrelación espacial y el diseño y análisis de experimentos. España: centro de Investigaciones Ambientales de Lourizán.

# MÉTODOS DE ESTIMACIÓN DE LA HETEROGENEIDAD ENTRE ESTUDIOS EN META-ANÁLISIS DE VARIABLES CONTINUAS

LIC. SOFÍA TAMUCH

Directora: MG. CRISTINA CUESTA

---

Una revisión sistemática es un documento de síntesis para identificar, evaluar e interpretar los resultados de estudios que abordan la misma pregunta de investigación. El meta-análisis consiste en la combinación de los resultados de estos estudios, con el propósito de llegar a una única conclusión, a partir de métodos estadísticos. Al combinar los resultados de los estudios, se pueden detectar efectos importantes de algún tratamiento, como así también se puede detectar una variabilidad entre estudios. El meta-análisis puede llevarse a cabo a través de dos metodologías: el modelo de efectos fijos, que considera que no hay heterogeneidad entre estudios y el modelo de efectos aleatorios, que sí la tiene en cuenta. Esta variabilidad adicional se modela utilizando un parámetro de variancia entre estudios y se han propuesto distintas estadísticas para estimarla. El objetivo de esta tesina es presentar métodos que permiten estimar la heterogeneidad entre estudios y sus propiedades. Se prueban estos métodos a partir de los resultados incluidos en una revisión sistemática actual sobre salud materna para el caso de medidas de resultados continuas. Los resultados muestran que los distintos métodos de estimación de la heterogeneidad afectan la estimación del efecto global en el caso en que la variabilidad sea alta o moderada.



## INTRODUCCIÓN

El gran crecimiento de publicaciones científicas ha sido un disparador para el desarrollo de la ciencia para la síntesis de la evidencia a lo largo de las últimas décadas. En este escenario, la identificación y lectura de todos los artículos que dan respuesta a una misma pregunta clínica es poco factible. Además, la apreciación crítica se dificulta especialmente cuando existe controversia acerca de las conclusiones de las publicaciones. Por ello, resulta indispensable disponer de estrategias eficaces para clasificar la evidencia válida de la que no lo es y para resumir la información existente. A partir de la necesidad de contar con una herramienta que permita sintetizar, valorar y poner al día toda la información médica existente, surgen las revisiones sistemáticas, que permiten identificar los estudios primarios que dan respuesta a una misma pregunta de interés. Si es posible, se unifican luego los resultados de los estudios seleccionados en la revisión sistemática a través de un meta-análisis.

El meta-análisis consiste en la combinación y resumen de los resultados de estudios independientes, con el propósito de llegar a una única conclusión, a partir de métodos estadísticos. Sin embargo, combinar los resultados de todos los estudios relevantes implica cuestionarse acerca de diferencias producidas por características propias de la población objetivo de cada estudio, por la definición o medición de las variables consideradas, como así también por la calidad metodológica derivada de cada diseño de investigación particular. Estas diferencias se manifiestan en la heterogeneidad de las variables estudiadas, la cual se define como una variación extra entre los resultados de los estudios además de la esperada por simple variación aleatoria.

En la práctica los métodos para combinar los resultados de los estudios dependen según se tenga en cuenta o no la heterogeneidad entre los mismos. Los modelos de efectos fijos consideran que no hay heterogeneidad entre estudios mientras que los modelos de efectos aleatorios sí la tienen en cuenta.

Los modelos de efectos aleatorios asumen que la variabilidad en las estimaciones del efecto (por ejemplo: razones de *odds* para variables dicotómicas o diferencias de medias para variables continuas) se puede descomponer en dos partes: variabilidad debida a la diferencia entre estudios y variabilidad debida a las respuestas dentro de los estudios.

Varios autores han propuesto metodologías para estimar la heterogeneidad entre estudios. La importancia de la determinación del método a utilizar radica en que la misma es relevante para estimar el efecto global a través de todos los estudios.

## **OBJETIVOS**

El objetivo principal de esta tesina es describir y comparar distintas metodologías para estimar la heterogeneidad entre estudios en meta-análisis de eventos continuos, a través de un ejemplo sobre la efectividad de diferentes tratamientos utilizados para prevenir la hemorragia postparto (HPP).

## **MATERIALES Y MÉTODOS**

Una revisión sistemática es un documento de síntesis que incluye la búsqueda y evaluación crítica, mediante métodos explícitos y reproducibles, de todos los estudios primarios que dan respuesta a una misma pregunta. Toda revisión sistemática tiene como finalidad minimizar los elementos de arbitrariedad mediante la descripción detallada del proceso de revisión de manera tal que, en principio, otra persona con acceso a los mismos recursos pueda realizarla y llegar, en general, a los mismos resultados.

Como punto de partida una revisión sistemática debe sustentarse en una pregunta clínica claramente definida. La pregunta clínica debe establecer: población bajo estudio, intervención que se quiere evaluar, comparación con otro/s tratamientos (puede haber o no) y el resultado de interés o lo que se quiere probar. En este sentido, el tipo de pregunta determina el diseño de estudio más idóneo para responder la pregunta. Por ejemplo, los ensayos controlados aleatorizados son el patrón de oro para responder preguntas sobre la efectividad de las intervenciones.

La revisión sistemática incluye la evaluación de la calidad de los estudios que cumplen con los criterios de selección y que constituyen la evidencia a ser resumida y analizada por los revisores. Si la revisión sistemática no se lleva a cabo siguiendo un protocolo, su rigor metodológico suele ser incierto y hay mayor probabilidad de introducción de sesgo (errores sistemáticos o desviaciones de la verdad que afectan la validez o confiabilidad de los

resultados). Siempre se intenta minimizar el sesgo durante todo el proceso de revisión, por ejemplo, al intentar establecer criterios de inclusión objetivos, de manera que los estudios potencialmente elegibles no se excluyan injustificadamente.

Una revisión sistemática puede resumirse en un meta-análisis, el cual combina los resultados de todos los estudios relevantes para obtener estimaciones más precisas de los efectos de una intervención. Las ventajas potenciales del meta-análisis incluyen el aumento de la potencia estadística y la exploración de las diferencias entre estudios. Puede ser realizado tanto cuando la variable de resultado es de tipo continua o bien categórica. La medida de efecto más utilizada para el primer caso es la diferencia de medias, siendo el riesgo relativo y la razón de *odds* las más empleadas cuando la variable respuesta es dicotómica. Se puede decir, entonces, que el meta-análisis es la integración de los resultados numéricos de los estudios primarios que, por medio del uso de técnicas estadísticas, permiten llegar a un estimador común del efecto bajo estudio (Green, 2011). Combinando los resultados de todos los estudios disponibles sobre un tema, se pueden detectar efectos importantes de algún tratamiento en estudio. Además del beneficio económico de evitar la realización de un nuevo estudio con un tamaño de muestra más grande.

El meta-análisis típicamente compara dos intervenciones, a las que se refieren generalmente como “experimental” y de “control”. En esta tesina el foco está puesto en comparar variables de resultado de tipo continua. Para cada estudio generalmente se informa la media, el desvío estándar y el tamaño de la muestra. Se consideran dos tipos de medidas de efecto diferentes para resultados continuos: diferencia de medias y diferencia de medias estandarizada. La diferencia de medias se utiliza generalmente cuando todos los estudios informan el resultado en la misma escala. Por otra parte, la diferencia de medias estandarizada se puede usar cuando los estudios miden el resultado en diferentes escalas.

Una vez que la diferencia de medias (estandarizada o no) y su variancia se han calculado, se continúa con los cálculos del meta-análisis con modelos de efectos fijos o de efectos aleatorios. El modelo de efectos fijos asume que los estudios que participan del meta-análisis provienen de una única población homogénea. Por lo tanto, para calcular una estimación general se promedian las estimaciones de cada estudio, teniendo en cuenta que algunas

estimaciones son más precisas que otras (las provenientes de estudios más grandes tiene estimaciones más precisas).

Sea  $\mu_{FE}$  (del inglés “*Fixed Effect*”) el efecto global de la intervención en la población (que se desea estimar). El modelo de efectos fijos que se postula es:  $\theta_k = \mu_{FE} + \varepsilon_k$ .

El “Método de Variancia Inversa” propone estimar el efecto global  $\mu_{FE}$  a partir del promedio ponderado de las estimaciones de los efectos individuales  $\theta_k$  con pesos  $w_k = 1/\hat{\sigma}_k^2$ . Bajo el supuesto que las estimaciones  $\theta_k$  son independientes, la variancia de  $\hat{\mu}_{FE}$  es igual al inverso de la sumatoria de los pesos.

El modelo de efectos aleatorios, en cambio, supone que cada estimación del efecto puede descomponerse en dos componentes de variancia mediante un proceso jerárquico de dos niveles. Se plantea el modelo de efectos aleatorios como:  $\theta_k = \mu_{RE} + \delta_k + \varepsilon_k$ .

Sea  $\delta_k$  la componente de variancia que indica el grado de heterogeneidad entre estudios y se describe como la diferencia entre la media  $\mu_{RE}$  (del inglés “*Random Effect*”) y el efecto específico de cada estudio, es decir, que el efecto global del tratamiento  $\mu_{RE}$  depende de él. Por último,  $\varepsilon_k$  es el término de error por el cual  $\theta_k$  difiere de  $\mu_{RE}$  y cuya variabilidad es  $\sigma_k^2$ . Por otro lado,  $\delta_k$  indica el efecto de los grupos y se considera aleatorio y con variancia  $\tau^2$ .

En este contexto el objetivo es estimar  $\mu_{RE}$  y  $\tau^2$  de manera óptima. Los criterios de optimalidad incluyen las propiedades de insesgamiento y eficiencia. El estimador de  $\mu_{RE}$  es el promedio ponderado de las estimaciones de los efectos individuales  $\theta_k$  con pesos  $w_k^* = 1/(\tau^2 + \sigma_k^2)$ . En la práctica,  $w_k^*$  debe sustituirse con los estimadores de los parámetros desconocidos de  $\tau^2$  y  $\sigma_k^2$ . Los parámetros del modelo generalmente se estiman bajo el supuesto de que las variaciones del estudio  $\sigma_k^2$  se conocen cuando en realidad se estiman, como consecuencia de calcular los valores de  $w_k^*$  utilizando estimaciones de  $\tau^2$  y  $\sigma_k^2$  surge lo siguiente. Es decir, la sustitución de las estimaciones insesgadas de  $\tau^2$  y  $\sigma_k^2$  en  $w_k^*$  da como resultado una estimación de la  $\widehat{Var}(\hat{\mu}_{RE})$  que está sesgada negativamente. Como consecuencia de este sesgo negativo, la variancia muestral de  $\hat{\mu}_{RE}$  se subestimarán en promedio.

Un segundo punto para considerar es el hecho que la variancia muestral de  $\hat{\mu}_{RE}$  es una función creciente de  $\tau^2$ . En consecuencia, la variancia muestral de  $\hat{\mu}_{RE}$  se subestimarán en promedio cuando se usa un estimador sesgado negativamente de  $\tau^2$ . Además, el parámetro  $\tau^2$  es de

interés no sólo para el cálculo de  $\hat{\mu}_{RE}$  y su variancia muestral, sino también porque indica directamente la heterogeneidad en los tamaños del efecto que no se puede explicar por error de muestreo.

Cualquier heterogeneidad en los efectos de la población, ya sea causada por interacción o por variaciones aleatorias, se reflejará en las estimaciones de  $\tau^2$  (mayor que cero).

Por otro lado, las estimaciones de  $\tau^2 = 0$  se interpretan generalmente como indicadores de la ausencia de interacción y de una variación aleatoria dentro de los efectos de la población, en cuyo caso, el modelo de efectos aleatorios se reduce al modelo simple de efectos fijos. Por lo tanto, la elección e interpretación del modelo apropiado está influenciada en gran medida por la estimación de  $\tau^2$ . Las estimaciones de  $\tau^2$  desempeñan un papel crucial en el meta-análisis. En resumen, se resalta la importancia de examinar las propiedades estadísticas de los diversos estimadores de heterogeneidad de la población porque: el uso de estimadores eficientes de  $\tau^2$  reduce la posibilidad de sobreestimar la precisión de  $\hat{\mu}_{RE}$ , el uso de estimadores sesgados de  $\tau^2$  lleva a una sobre o subestimación de la precisión de  $\hat{\mu}_{RE}$  y las estimaciones de  $\tau^2$  son relevantes para la elección del modelo y el análisis de posibles interacciones.

Los estimadores de heterogeneidad permiten obtener una medida global del efecto en el modelo de efectos aleatorios, de forma similar al método de Variancia Inversa que aplica el modelo de efectos fijos, sólo que el peso de cada estudio varía para incorporar tanto la variabilidad intra-estudio como la variabilidad entre-estudios.

Algunos de los estimadores de heterogeneidad entre estudios ( $\tau^2$ ) son el estimador de DerSimonian-Laird (DL), estimador de Hunter y Schmidt (HS), el estimador propuesto por Hedges (EH) y dos estimadores basados en la estimación de máxima verosimilitud, el estimador de máxima verosimilitud (MV) y el estimador de máxima verosimilitud restringida (MVR).

Las diferencias entre los estimadores pueden explicarse en cierta medida por los diferentes esquemas de ponderación que utilizan. La mayoría de los métodos incluyen tanto las variancias dentro del estudio como una estimación de la variancia entre los estudios en las ponderaciones, el método de DL utiliza sólo la inversa de las variancias dentro del estudio, mientras que el estimador Hedges usa igual peso para todos los ensayos. El estimador de

Máxima Verosimilitud tiende a subestimar la heterogeneidad de la población al no tener en cuenta el hecho, de que  $\mu_{RE}$  en la ecuación del estimador también se estima a partir de los datos, en cambio, el estimador de Máxima Verosimilitud Restringida compensa esta subestimación mediante el uso de una combinación lineal del vector de respuestas, de manera que los datos transformados estén libres de los efectos fijos. Una desventaja de los estimadores iterativos es que dependen de la elección del método de maximización, que podría no converger en una solución y, por lo tanto, el estimador no proporciona un valor de  $\tau^2$  (Veroniki, 2015).

## RESULTADOS

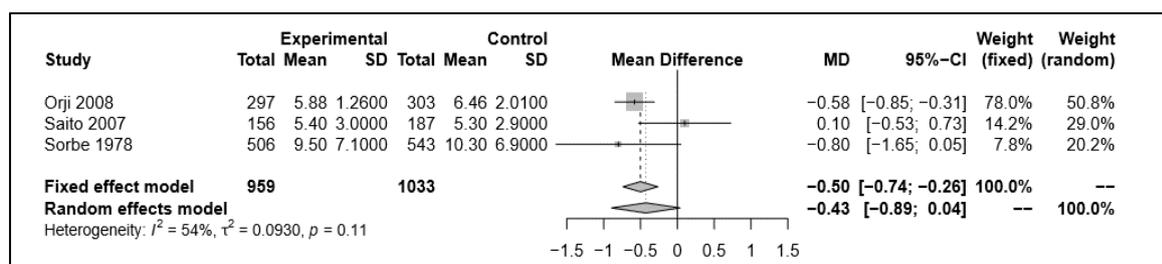
En esta tesina, se muestra un ejemplo relativo a la efectividad de diferentes tratamientos utilizados para prevenir la hemorragia postparto (HPP), que es una complicación común y ocasionalmente mortal del trabajo de parto. La hemorragia postparto se define como la pérdida de más de 500 ml de sangre durante o después del alumbramiento. El alumbramiento, también llamado tercera etapa del parto, es el período desde el nacimiento del bebé hasta la expulsión de la placenta. La mayoría de las mujeres reciben medicamentos que estimulan directamente el útero (uterotónicos profilácticos) durante el parto para prevenir las hemorragias que resultan de la falla del músculo uterino para contraerse normalmente. Sin embargo, no hay consenso aún acerca del mejor tratamiento que puede emplearse, de allí el surgimiento de las revisiones sistemáticas sobre este tema y la necesidad de combinar sus resultados a través de meta-análisis.

El ejemplo de aplicación proviene de la revisión sistemática “Oxitocina profiláctica en la tercera etapa del parto para prevenir la hemorragia postparto” (Westhoff G, Cotter AM, Tolosa JE, 2018). La misma es una Revisión Cochrane publicada en el año 2013. Estas revisiones siguen los lineamientos propuestos por el grupo de colaboración Cochrane, el cual es una red sin fines de lucro de más de 13.000 miembros en todo el mundo y cuya misión es promover la evidencia científica para los tomadores de decisión en salud produciendo revisiones sistemáticas accesibles, relevantes y de alta calidad (Cochrane, 2019).

El objetivo de la revisión fue determinar la efectividad de la oxitocina profiláctica en cualquier dosis para prevenir la HPP. Se incluyeron 20 ensayos controlados aleatorizados (en total 10.806 mujeres embarazadas que anticipaban un parto natural). Se les administró oxitocina profiláctica a 4.203 mujeres, un medicamento que ayuda a que el útero se contraiga y se reduzca la cantidad de pérdida de sangre durante el alumbramiento.

Se comparó la oxitocina contra los alcaloides, que son fármacos uterotónicos que reducen la cantidad de sangrado y la necesidad de transfusión sanguínea. La intervención de la revisión que se tuvo en cuenta para el meta-análisis fue: el uso de oxitocina profiláctica para medir la duración del alumbramiento (período desde el nacimiento del bebé hasta la expulsión de la placenta en segundos).

Figura 1. Oxitocina vs. alcaloides en la duración media del alumbramiento (en seg.)



Fuente propia. Salida elaborada con librería “meta” de R.

El objetivo del meta-análisis mostrado en la Figura 1 es medir el efecto de la oxitocina profiláctica contra los alcaloides siendo en este caso la variable respuesta la duración media del alumbramiento (en segundos). Entre los 20 estudios iniciales, se identificaron 3 estudios que realizan esta comparación.

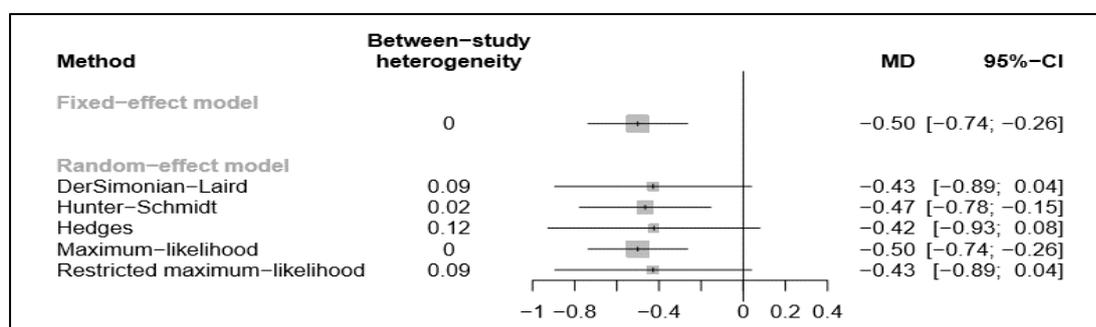
La particularidad de este caso es el resultado de la prueba de heterogeneidad, que no resulta significativo, aún con un grado de heterogeneidad del 54%, es decir, heterogeneidad moderada. En términos del problema esto significa que la magnitud del efecto de la oxitocina profiláctica en la duración media de la tercera etapa del parto con respecto al placebo puede considerarse similar en los distintos estudios.

La estimación de la diferencia de medias en el modelo de efectos fijos fue -0.50 segundos, esto quiere decir, que la duración promedio del alumbramiento fue menor con la oxitocina

que con alcaloides, con un intervalo de confianza del 95% de (-0.74 ; -0.26), en cambio, en el modelo de efectos aleatorios la duración promedio del alumbramiento fue menor con la oxitocina que con alcaloides con -0.43 segundos y un intervalo de (-0.89 ; 0.04).

Finalmente se opta por el modelo de efectos aleatorios, aunque la prueba de heterogeneidad no sea significativa, ya que la heterogeneidad entre estudios es moderada. Esta elección implica optar por una estimación más conservadora del efecto del tratamiento.

Figura 2. Comparación de los métodos de estimación para la heterogeneidad entre estudios



Fuente propia. Salida elaborada con librería “meta” de R.

La diferencia no es tan notoria entre la estimación puntual del modelo de efectos fijos (que no tiene en cuenta la heterogeneidad entre estudios) y las estimaciones en el modelo de efectos aleatorios. Esto se debe a que si bien hay una heterogeneidad moderada entre los estudios la misma no es significativa (Figura 2).

En cuanto a los valores de los distintos métodos de estimación de la heterogeneidad entre estudios en el modelo de efectos aleatorios, los estimadores de HS y MV producen los valores más bajos comparado con los otros métodos. Una desventaja de los estimadores iterativos, como el de MV, es que dependen de la elección del método de maximización, que podría no converger en una solución y, por lo tanto, el estimador no proporciona un valor de  $\hat{\tau}^2$  (Veroniki, 2015). Se destaca también que, tanto el estimador de DL como el de MVR producen valores iguales de  $\hat{\tau}^2$ . Por último, el estimador de Hedges es el que en este caso sobreestima la heterogeneidad de la población, esto principalmente ocurre cuando la cantidad de estudios es pequeña.

## CONCLUSIONES

En la actualidad son abundantes los artículos que describen resultados de meta-análisis en publicaciones médicas. La razón fundamental para hacer un meta-análisis es aumentar la potencia y la precisión a la hora de valorar los efectos de un tratamiento. Otra ventaja es poder evaluar conjuntamente los resultados de múltiples publicaciones, además del beneficio económico que llevaría la realización de un nuevo estudio con un tamaño de muestra más grande.

En el modelo de efectos fijos, los efectos observados del tratamiento se distribuyen alrededor de un efecto común, con una variación que depende completamente de las diferencias dentro del estudio. El modelo de efectos aleatorios, en cambio, supone que cada estimación del efecto puede descomponerse en dos componentes de variancia. Por lo tanto, la variabilidad se atribuye tanto a la variancia dentro del estudio, debido al error de muestreo, como a la variancia entre estudios. En consecuencia, la variancia estimada entre estudios influye en las ponderaciones asignadas a cada estudio en un meta-análisis y, por lo tanto, en su precisión.

Las diferencias entre los estimadores de la heterogeneidad entre estudios pueden explicarse en cierta medida por los diferentes esquemas de ponderación que utilizan. Los resultados de los estudios disponibles dependen de los escenarios particulares que investigaron y de las propiedades de los métodos que se examinaron. Se requiere investigación adicional para evaluar las propiedades de todos los métodos bajo los mismos escenarios a través de un estudio de simulación integral, teniendo en cuenta la estimación por intervalo de confianza de los métodos de estimación de heterogeneidad entre estudios.

## BIBLIOGRAFÍA

*Cochrane*. (Febrero de 2019). Obtenido de <https://es.cochrane.org/es/revisiones-cochrane>

Green, J. H. (2011). *Manual Cochrane de Revisiones Sistemáticas de Intervenciones*. Obtenido de [https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/manual\\_cochrane\\_510\\_web.pdf](https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/manual_cochrane_510_web.pdf)

Guido Schwarzer, G. R. (2015). *Meta-Analysis with R*. Springer.

Veroniki, A. A. (2015). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Wiley Online Library*, 55-79.

Vietchtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random Effects Model. *Journal of Educational and Behavioral Statistics Fall 2005, Vol. 30.*

Westhoff G, C. A. (2018). *Prophylactic oxytocin for the third stage of labour to prevent postpartum haemorrhage (Review)*. Obtenido de <https://www.cochranelibrary.com/>

Zwinderman, T. J. (2017). *Modern Meta-Analysis*. Springer.

# VECTORES DE CAPACIDAD DE PROCESOS MULTIVARIADOS. UN ESTUDIO COMPARATIVO

LIC. LUISINA VAZQUEZ

Directora: DRA. DANIELA DIANDA

Codirectora: DRA. MARTA QUAGLINO

El Análisis de Capacidad de Procesos es una de las técnicas más utilizadas durante la implementación de estrategias de mejora de la calidad. Su enfoque univariado ha tenido un amplio tratamiento, pero su contraparte multivariada está aún en pleno desarrollo. En ese contexto se sitúa la temática de este trabajo, ahondando en uno de los enfoques propuestos en la literatura para analizar la capacidad de un proceso multivariado: los vectores de capacidad de procesos multivariados.

Si bien los primeros intentos de cuantificación de la capacidad de un proceso multivariado se plasmaron como medidas unidimensionales, que en cierto sentido pueden verse como generalizaciones de las propuestas univariadas más conocidas, en muchas situaciones, estos valores únicos impiden o dificultan al usuario conocer la naturaleza de la falta de capacidad, cuando ésta ocurre. Como alternativa a estas medidas unidimensionales, se proponen los denominados vectores de capacidad: medidas multidimensionales en las que cada componente refleja el comportamiento del proceso en un único aspecto. En este trabajo se presentan cuatro de los vectores de capacidad existentes en la literatura y se lleva a cabo un estudio comparativo de estas alternativas para evaluar la performance de las mismas ante diferentes particularidades de los procesos.



## INTRODUCCIÓN

La demanda de calidad siempre ha formado parte de la naturaleza humana, sin embargo, el intento de cuantificar y establecer estándares formales de calidad es indudablemente un fenómeno del siglo XX. El concepto de calidad ha evolucionado a lo largo de los años, especialmente en las últimas décadas, cuando se asumió como prioridad en el campo empresarial y se convirtió en la premisa principal para las organizaciones.

Dependiendo del objetivo que se persiga, pueden encontrarse variadas concepciones del término calidad. Sin embargo, en todas ellas está presente la filosofía de “adecuación a los requerimientos” impuestos por el mercado. La técnica abordada en esta tesina, el análisis de capacidad de procesos, tiene como propósito justamente medir esa adecuación, la cercanía entre el resultado del proceso y lo que se espera de él, en términos de características identificables y medibles. Específicamente, esta técnica compara las características de los productos resultantes del proceso con las características esperadas para los mismos, llamadas requerimientos o especificaciones y resume dicha comparación en términos de cantidades conocidas como índices de capacidad, las cuales permiten determinar cuantitativamente si un proceso es capaz, es decir, si logra satisfacer las especificaciones.

El desarrollo inicial de esta técnica se dio en el contexto univariado, es decir asumiendo que el proceso bajo estudio posee una única característica de interés a ser evaluada. Sin embargo, la situación más realista y habitual en los procesos actuales, se da en el contexto multivariado, donde existen dos o más características que describen el resultado y que deben ser evaluadas simultáneamente para poder determinar la capacidad global del proceso. Surgió así el análisis de capacidad multivariado, que ha sido objeto de estudio durante los últimos años, en los que muchos autores desarrollaron y propusieron diferentes alternativas para los índices multivariados. No obstante, este tema se encuentra aún en desarrollo y hay muchos aspectos no resueltos en torno a su aplicación en contextos prácticos. Uno de tales aspectos se relaciona con la dimensión de la medida de resumen que se utiliza para cuantificar la capacidad de un proceso multivariado. Los vectores de capacidad son indicadores multidimensionales del comportamiento del proceso, cuyas componentes están diseñadas para monitorear cada una, un único aspecto del mismo (variabilidad, centrado, proporción de productos no conformes, etc.). Se evalúa la performance de algunas medidas multidimensionales mediante su aplicación en diferentes escenarios.

## ANÁLISIS DE CAPACIDAD MULTIVARIADO

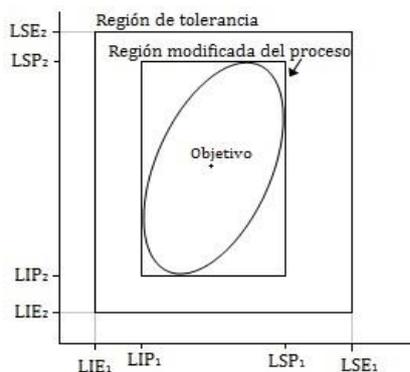
El análisis de capacidad tiene como propósito evaluar el comportamiento de un proceso en relación a los requerimientos impuestos sobre los productos y resumir cuantitativamente el resultado de dicha comparación. En el contexto multivariado, la capacidad de un proceso depende de varias variables de calidad ( $v$ ). Así, el comportamiento del proceso podría representarse a través de la región en el espacio  $v$ -dimensional en el que el proceso genera resultados. Si para las variables de calidad se asume una distribución conjunta normal, entonces dichas regiones, llamadas regiones de variación natural del proceso, estarán determinadas por hiperelipsoides de densidad constante.

Por otro lado, si las especificaciones para cada una de las variables de calidad, vienen dadas como valores discretos independientes para cada una de ellas, es decir, un límite inferior de

especificación ( $LIE$ ) y un límite superior de especificación ( $LSE$ ), entonces puede definirse la región de especificación como el hiperrectángulo generado por el producto cartesiano de los intervalos de especificación individuales. El análisis de la capacidad del proceso consiste en comparar la región de variación natural del proceso, que es la región hiperelipsoidal en la que el proceso realmente opera, con la región de especificación, que es la región hiperrectangular donde se permite que el proceso opere. Todas las propuestas para llevar a cabo dicha comparación, consideran una corrección de forma en alguna de las dos regiones, de modo que la comparación no se vea afectada por la diferencia en las formas geométricas de ambas regiones. En este trabajo se presentan cuatro alternativas de vectores de capacidad multivariados.

Hubele *et al.* (1991) realizaron uno de los primeros aportes en esta línea, trabajando sobre procesos bivariados. Su propuesta consiste en representar la capacidad de un proceso mediante un vector de tres componentes:  $VCH = (C_{pM}; PV; M)$ . La primera componente del vector ( $C_{pM}$ ) es un índice que compara los volúmenes de las regiones del proceso y de especificación, modificando previamente la región del proceso al rectángulo más pequeño que circunscribe a la elipse de probabilidad  $1 - \alpha$  (Figura 1).

Figura 1. Ejemplo de una región de proceso modificada según la propuesta de Hubele *et al.*



Fuente: Adaptado de Hubele *et al.* (1991).

Las aristas de esta nueva región rectangular se obtienen resolviendo un problema de extremos condicionados y el índice se define como el cociente entre el área del rectángulo de especificación y el del rectángulo de variación del proceso. Así, valores mayores a uno señalan que la dispersión del proceso es menor a la permitida por especificación, mientras que valores inferiores señalan falta de capacidad del proceso por excesiva variabilidad.

La segunda componente del vector ( $PV$ ), evalúa la capacidad del proceso en términos de centrado, a través de la probabilidad asociada a la estadística del test  $T^2$  de Hotelling que prueba la igualdad entre el vector de valores objetivos y el vector de medias del proceso. De este modo, valores superiores al nivel de significación que se haya establecido, señalan procesos capaces en centrado, mientras que valores pequeños indican que el proceso no opera de acuerdo a los valores nominales especificados.

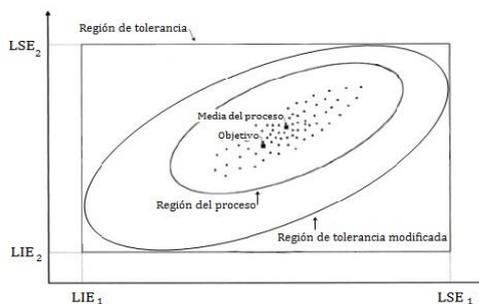
Como tercera componente, proponen una medida que refleja la localización del proceso respecto de los límites de especificación ( $M$ ), de modo que si se obtiene un valor igual a 1 el proceso está

contenido completamente dentro del rectángulo de especificación y, en caso que el proceso rebese la zona de especificación, se obtiene un valor mayor que 1.

Años más tarde, Shahriari *et al.* (1995) extienden la propuesta de Hubele *et al.* (1991) para considerar procesos con más de dos variables de calidad, denominando al nuevo vector de capacidad como  $MCPV = [C_{PM}, PV, LI]$ . La extensión modifica sólo la primera componente del vector, la cual se define como un cociente ya no de áreas, sino de volúmenes de regiones hiperrectangulares. Además, se propone tener en cuenta el número de dimensiones del proceso tomando la raíz  $v$ -ésima de dicho cociente. Las restantes dos componentes se mantienen iguales a las propuestas por Hubele *et al.* (1991).

En el año 2009 surge otra propuesta de vectores de capacidad, también de tres componentes, debida a los autores Shahriari y Abdollahzadeh, quienes lo denominaron  $NMPCV = [NMC_{pm}^*, PV, LI]$ . En este vector, la segunda y tercera componentes son similares a las de los vectores ya definidos, pero la primera componente compara especificaciones y variación natural del proceso modificando la forma de la región de especificación, de modo que las regiones a comparar sean hiperelipsoidales. Para ello, proponen tomar como región de especificación modificada al hiperelipsoide más grande centrado en el objetivo, que se encuentre completamente dentro de la región de tolerancia original y cuyos ejes sean paralelos a los del elipsoide del proceso (Figura 2). El cociente de volúmenes resultante se interpreta de manera análoga a los casos anteriores.

Figura 2. Ejemplo de una región de tolerancia modificada según la propuesta de Shahriari y Abdollahzadeh para  $v = 2$

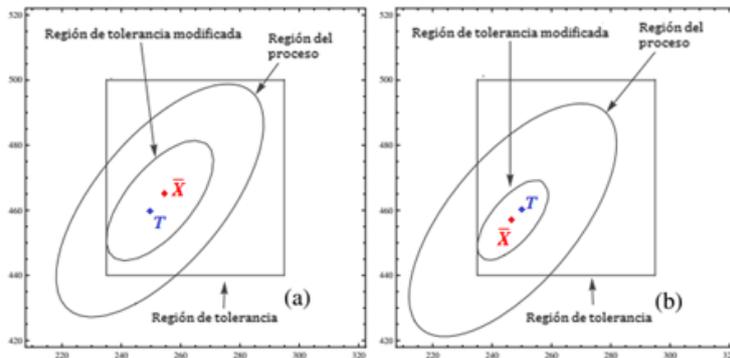


Fuente: Adaptado de Shahriari y Abdollahzadeh (2009).

La última propuesta analizada es bastante reciente y fue diseñada específicamente para el caso de procesos con especificaciones asimétricas, es decir, procesos en los que, para al menos una de las variables de calidad, su valor objetivo no coincide con el punto medio del intervalo de especificación. Cuando esto sucede, las desviaciones que se producen en la dirección opuesta a la de la asimetría son más graves que las que se producen en el otro sentido y ello debería ser tenido en cuenta en la evaluación de la capacidad del proceso. Es por esto que los autores Ganji y Gildeh (2015) introdujeron un nuevo vector de capacidad multivariado, en el que una de sus componentes intenta, en caso de existir un corrimiento en el centro del proceso, capturar no sólo su presencia sino también su sentido. En esencia, este índice es similar al propuesto por Shahriari y Abdollahzadeh, pero la diferencia radica, nuevamente, en la forma en la que se modifica la región de tolerancia. Para la modificación, los autores proponen considerar el hiperelipsoide más grande con ejes paralelos a los ejes del elipsoide original del proceso, que se

encuentra completamente contenido en la región de tolerancia original, pero centrado en el vector de medias y no en el vector de valores objetivo (Figura 3).

Figura 3. Ejemplo de regiones de tolerancia modificadas según la propuesta de Ganji y Gildeh (2015) para  $\nu = 2$



Fuente: elaboración propia.

Así, si el proceso tiene un corrimiento en el mismo sentido en que se presenta la asimetría de las tolerancias (Figura 3, panel a), la región de tolerancia resultante es más grande que si el corrimiento se da en sentido opuesto al de la asimetría (Figura 3, panel b). De este modo, el índice de estos autores penaliza distinto según el sentido del corrimiento del proceso.

El vector propuesto consiste en dos componentes:  $MPCVG = [MC_{pG}, PV]$ . La primera es el cociente de los volúmenes de las regiones mencionadas, con interpretación análoga a la de los índices anteriores y la segunda componente,  $PV$  coincide con las propuestas anteriores.

## ESTUDIO COMPARATIVO

La utilidad y las limitaciones de los índices tratados hasta aquí son importantes para los usuarios, por lo que se realizó un estudio comparativo que permitiera evaluar su performance frente a distintos escenarios.

Las situaciones prácticas que se plantearon, consideran sólo dos variables de calidad ( $\nu = 2$ ) a fin de facilitar la visualización gráfica de las particularidades en su definición.

Como primer paso en la ejecución de este estudio comparativo, fue necesario desarrollar en software estadístico los programas que permitan calcular cada uno de los vectores de capacidad a analizar, ya que hasta el momento sólo el vector de Shahriari *et al.* (1995) se encuentra implementado y disponible en el software libre R.

El estudio comparativo planteó procesos con diferentes estados de capacidad, de modo de poder evaluar si las distintas propuestas lograban detectar correctamente la situación del proceso. Esto se combinó también con variantes en el tipo de especificaciones impuestas (simétricas o asimétricas), generando un total de 7 escenarios diferentes:

- Caso 1: Proceso capaz, con especificaciones simétricas.
- Caso 2: Proceso no capaz en dispersión, con especificaciones simétricas.
- Caso 3: Proceso no capaz en media, con especificaciones simétricas.
- Caso 4: Proceso no capaz en media ni en dispersión, con especificaciones simétricas.
- Caso 5: Proceso capaz, con especificaciones asimétricas.
- Caso 6: Proceso no capaz en media, con especificaciones asimétricas y desplazamiento del centro del proceso coincidente con la asimetría.
- Caso 7: Proceso no capaz en media, con especificaciones asimétricas y desplazamiento del centro del proceso opuesto a la asimetría.

Cada uno de los casos propuestos fue generado mediante la definición de valores objetivos, límites de especificación, vectores de medias y matrices de variancias y covariancias acordes a la situación que se intenta representar (Tabla 1). En cada escenario, fue simulada una muestra de 100 observaciones proveniente de una distribución normal bivariada con los parámetros especificados para cada uno, con la cual se calcularon los cuatro vectores de capacidad a comparar.

Tabla 1. Parámetros y especificaciones de los procesos simulados bajo cada escenario

Caso	Vector de medias “ $\mu$ ”	Matriz de Var. y Covar. “ $\Sigma$ ”	Vector target “ $T$ ”	Límites inferiores de especificación “LIE”	Límites superiores de especificación “LSE”
1	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 1 & 0,56 \\ 0,56 & 0,64 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
2	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 1,44 & 1,176 \\ 1,176 & 1,96 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
3	$\begin{pmatrix} 13 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0,56 \\ 0,56 & 0,64 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
4	$\begin{pmatrix} 13 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 1,44 & 1,176 \\ 1,176 & 1,96 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
5	$\begin{pmatrix} 11 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 0,36 & 0,21 \\ 0,21 & 0,25 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
6	$\begin{pmatrix} 12 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 0,36 & 0,21 \\ 0,21 & 0,25 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$
7	$\begin{pmatrix} 10 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 0,36 & 0,21 \\ 0,21 & 0,25 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 10 \end{pmatrix}$

Fuente: elaboración propia.

## RESULTADOS

La Tabla 2 muestra los resultados de los cuatro vectores de capacidad analizados, obtenidos bajo cada uno de los escenarios hipotéticos planteados. Las celdas de dicha tabla fueron coloreadas para brindar una rápida apreciación de la performance de los vectores bajo cada situación. Así, el color verde señala las situaciones en las que el vector logró determinar correctamente el estado real de capacidad del proceso, mientras que el rojo indica lo contrario y

el amarillo señala situaciones en las que el vector podría tener problemas para realizar un análisis correcto del proceso.

Tabla 2. Resultados obtenidos para los cuatro vectores propuestos bajo cada escenario

Caso	VCH=( $C_{pM}$ PV M)	MCPV=( $C_{pM}$ PV LI)	NMPCV=( $NMC_{pM}^*$ PV LI)	MPCVG=( $MC_{pG}$ PV)
1	(1,27 0,2 1)	(1,13 0,2 1)	(1,08 0,2 1)	(1,03 0,2)
2	(0,62 0,2 1,36)	(0,79 0,2 0)	(0,61 0,2 0)	(0,56 0,2)
3	(1,28 0 1,11)	(1,13 0 0)	(1,08 0 0)	(0,76 0)
4	(0,62 0 1,52)	(0,79 0 0)	(0,61 0 0)	(0,36 0)
5	(3,41 0,2 1)	(1,85 0,2 1)	(1,48 0,2 1)	(1,53 0,2)
6	(3,41 0 1)	(1,85 0 1)	(1,48 0 1)	(1,68 0)
7	(3,41 0 1)	(1,85 0 1)	(1,48 0 1)	(1,04 0)

Fuente: elaboración propia.

En los casos 1, 2 y 5 de la Tabla 2, los cuatro vectores lograron identificar correctamente el estado de capacidad real del proceso. Para los casos 3 y 4, los tres primeros vectores identificaron correctamente la situación real del proceso, pero, el vector de Ganji y Gildeh generó confusión. Esto se debe, a que se obtuvo un valor menor que 1 en la primera componente y por la forma en que se construye el índice no se puede concluir si la falta de capacidad se debe a exceso de variabilidad o a problemas de centrado. Para los casos 6 y 7, a pesar de que se plantea un desplazamiento de igual magnitud en ambos, en el caso 6, el desplazamiento tiene menos posibilidad de generar un problema de capacidad. La razón es que el corrimiento se da en el mismo sentido de la asimetría de las tolerancias. El desplazamiento del caso 7 es más grave, ya que genera que la región del proceso quede localizada en el borde de la región de especificación. No obstante, los tres primeros vectores de capacidad arrojan exactamente los mismos valores en ambas situaciones, con los cuales se concluye que el proceso es capaz en dispersión, pero no en localización. La única alternativa que toma en cuenta esta particularidad en el comportamiento de ambos procesos es la propuesta de Ganji y Gildeh (2015), penalizando el corrimiento menos favorable con una disminución en el valor del índice de capacidad (primera componente del vector). El índice resulta 1,04 en el caso 7 y 1,60 para el caso 6. Si bien ambos valores son superiores a la unidad, la diferencia en magnitud permite diferenciar ambas situaciones.

## COMENTARIOS FINALES

En este trabajo se ha abordado la temática concerniente al análisis de capacidad multivariado, específicamente en la línea que propone la creación de medidas multidimensionales, denominadas vectores de capacidad. La idea de crear un indicador multidimensional surge como respuesta a la dificultad que presentan las medidas unidimensionales para resumir en un único valor todos los aspectos involucrados en un proceso multivariado y alertar al usuario sobre la causa de la falta de capacidad, en caso de ocurrir. Así, los vectores de capacidad fueron diseñados con la idea de que cada componente relate sobre un aspecto del proceso en particular.

Se presentan cuatro propuestas para la creación de vectores de capacidad, analizando sus particularidades y evaluando su performance a través de un estudio comparativo bajo diferentes escenarios hipotéticos. Los resultados encontrados permiten derivar algunos patrones de comportamiento de las diferentes propuestas. En principio, se observa que no existe una propuesta con comportamiento superador bajo todas las situaciones. Los vectores  $VCH$ ,  $MCPV$  y  $NMPCV$  logran identificar correctamente el estado de capacidad real de los procesos a través de todas las situaciones. Sin embargo, en caso de tener especificaciones asimétricas, estas medidas no tienen en cuenta que los desplazamientos del proceso en diferentes direcciones pueden no ser igualmente nocivas para el proceso. Para dichas situaciones, el vector  $MPCVG$  muestra mejor comportamiento, penalizando de modo diferente los corrimientos del proceso según la dirección en la que se producen. No obstante, en otras situaciones este vector no logra emitir un juicio claro sobre el comportamiento del proceso.

Por todo lo expuesto, se puede concluir que la elección de una u otra alternativa depende de las características del problema particular que se esté tratando. Si bien los ejemplos de aplicación propuestos en el estudio comparativo son particulares, permitieron descubrir y evidenciar ciertas particularidades en su adecuación.

## BIBLIOGRAFÍA

- Dianda, D. (2015). Estudio Estadístico de Sistemas de Medida e Indicadores de Capacidad de Procesos Multivariados, en Contextos de Mejora de la Calidad y la Productividad. Tesis Doctoral. Universidad Nacional de Rosario.
- Ganji, Z.A. y Gildeh, B.S. (2015). A modified multivariate process capability vector. *International Journal of Advanced Manufacturing Technology*, 83: 1221–1229.
- Hubele, N.F., Shahriari, H., Cheng, C.S. (1991). A bivariate process capability vector, in statistics and design in process control, in *Statistical Process Control in Manufacturing* edited by Keats, J.B. y Montgomery, D.C. 299-310. ASQC Quality Press.
- Kane, V.E. (1986). Process capability indices. *Journal of Quality Technology*, 18: 41-52.
- Kotz, S. y Johnson, N.L. (1993). *Process capability indices*. Chapman and Hall.
- Montgomery, D.M. (1991). *Control estadístico de la calidad*. Grupo Editorial Iberoamérica.
- Pearn, W.L. y Kotz S. (2006). *Encyclopedia and Handbook of Process Capability Indices: A Comprehensive Exposition of Quality Control Measures*. World Scientific Publishing Co. Pte. Ltd.
- Santos Fernández, E., Scagliarini, M. (2012). MPCV: An R package for computing multivariate process capability indices. *Journal of Statistical Software*, 47: 1-15.

Shahriari, H., Hubele, N.F., Lawrence, F.P. (1995). A multivariate process capability vector. *Proceedings of the 4th Industrial Engineering Research Conference*, Institute of Industrial Engineers, 304-309.

Shahriari, H. y Abdollahzadeh, M. (2009). A new multivariate process capability vector. *Quality Engineering*, 21(3): 290-299.

Taam, W., Subbaiah, P. y Liddy, J.W. (1993). A note on multivariate capability indices. *Journal of Applied Statistics*, 20(3): 339-351.

Wang, F., Hubele, N., Lawrence, F., Miskulin, J. y Shahriari, H. (2000). Comparison of three multivariate process capability indices. *Journal of Quality Technology*, 32(3): 263–275.



CONSEJO PROFESIONAL  
DE CIENCIAS ECONOMICAS  
DE LA PROVINCIA DE SANTA FE



Colegio de Graduados  
en Ciencias Económicas  
de Rosario