

## **Análisis de Datos Conteo con enfoque modelado: Cantidad de dolencias musculoesqueléticas y factores asociados en profesionales de enfermería**

### **Modelling approach to Count data analysis: Number of musculoskeletal conditions and associated factors in nursing professionals**

Fernando Gimenez Sena<sup>1</sup> 

<sup>1</sup>Universidad Nacional de Asunción, Facultad de Ciencias Exactas y Naturales, Departamento de Estadística. Email: [ferdisena@facen.una.py](mailto:ferdisena@facen.una.py).

**Resumen:** Las dolencias derivadas de la actividad laboral son causa de preocupación creciente. Los trabajadores de enfermería son particularmente propensos a sufrir molestias musculoesqueléticas. Con los datos de un estudio transversal con 345 individuos, profesionales de enfermería del Hospital de Clínicas en San Lorenzo, utilizando modelos para datos conteos, se investigan factores asociados con este tipo de dolencia. Se encuentra que efectivamente existe asociación entre el tipo de calzado y la cantidad de patologías; pero al incorporar factores individuales, como sexo y edad, y factores laborales, como antigüedad y cantidad de turnos atendidos, el efecto del tipo de calzado utilizado se diluye.

**Palabras clave:** Datos conteo; modelos Poisson, Binomial negativo, Hurdle; sobredispersión; trastorno musculoesquelético.

**Abstract:** Work related ailments are a cause of growing concern. Nursing workers are particularly prone to musculoskeletal conditions. A study at the Hospital de Clínicas in San Lorenzo found an association between the type of work shoes used and the number of pathologies per individual. With the same data, from a cross-sectional study of 345 units, using models for count data, other factors that could be associated are investigated. It is found that there is indeed an association between the type of footwear and the number of pathologies. But, by incorporating individual factors, such as sex and age, and employment factors, such as seniority and number of shifts attended, the effect of footwear type dilutes completely.

**Keywords:** Count data; Poisson, Binomial negative, Hurdle models; overdispersion; musculoskeletal disorder.

## **Introducción**

A medida que las exigencias laborales se hacen más intensas también aumentan los problemas físicos y psicológicos asociados con estas actividades. Según la Organización Internacional del trabajo (OIT), alrededor de 160 millones de personas sufren enfermedades relacionadas con el trabajo, citado por Zamora *et al.* (2019). Por este motivo, las iniciativas para desarrollar diversos aspectos de la seguridad laboral han ido adquiriendo mayor importancia en el transcurso del tiempo.

En el caso de los trabajadores de la salud, concretamente los que ejercen como profesionales de enfermería, el continuado contacto con pacientes y familiares aquejados con dolencias de diversa gravedad, las prolongadas horas de servicio y las condiciones desfavorables propician el deterioro de la salud física y mental, Muñoz *et al.* (2015); como

caso extremo, se ha constatado esta realidad durante la actual pandemia del COVID.

Los problemas de salud de las enfermeras pueden categorizarse en a) riesgos biológicos, b) riesgos químicos, c) riesgos físicos, y d) riesgos psicológicos (Tan, 1991). Estos riesgos generan problemas de salud en los trabajadores y por extensión generan disminución de la productividad, la calidad y seguridad de la atención a los pacientes. Letvak & Ruhm, (2011).

En un estudio de Trinkoff *et al.* (2002), los participantes de su estudio han presentado síntomas relacionado con afecciones musculoesqueléticas con al menos una semana de duración, al menos una vez al mes durante el año anterior a su entrevista.

Según la Organización Mundial de la Salud (OMS), Luttmann *et al.* (2003), se define como trastornos musculoesqueléticos (TME) a proble-

Recibido: 19/12/2021 Aceptado: 17/08/2022



mas de salud del aparato locomotor que abarcan: músculos, tendones, esqueleto óseo, cartílagos, ligamentos y nervios. Los individuos pueden presentar desde molestias leves y pasajeras hasta lesiones irreversibles y discapacitantes, muchas de ellas causadas o intensificadas por el trabajo. Así mismo, la OMS señala que son factores que influyen en el desarrollo de TME, ejercer demasiada fuerza, manipulación manual de cargas por periodos prolongados, manipulación de objetos de manera repetida y frecuente, trabajar en posturas perjudiciales, esfuerzo muscular estático, inactividad muscular, movimientos repetitivos, exposición a vibraciones, factores ambientales, riesgos físicos y factores psicosociales.

En particular, Anderson & Nester, (2017) relacionan problemas en las extremidades inferiores y la espalda con las tareas ocupacionales, los pies, el calzado y la superficie del recinto laboral. Consideraciones similares se encuentra en Reed *et al.* (2014), quienes identifican factores relacionados con el individuo (edad, índice de masa corporal, condiciones físicas generales, etc.), características del calzado laboral, y factores relacionados con el lugar de trabajo (nivel profesional, lugar de trabajo, horas laborales trabajadas promedio, etc.).

En Maidana *et al.* (2020), se menciona que un calzado inadecuado potencia lesiones en los pies y las uñas (el aparato ungueal, que pueden producir bajas laborales; estas lesiones pueden revestir diversa gravedad; y aún los casos leves son potencialmente crónicas o y podrían desembocar en patologías más complejas. Estos autores, en su estudio encuentran evidencias de asociación entre el tipo de calzado utilizado y la cantidad de trastornos musculoesqueléticos reportados por los trabajadores de enfermería del Hospital de Clínicas en San Lorenzo (HCSL).

En este trabajo se exploran otros factores, además del tipo de calzado, registrados en Maidana *et al.* (2020) que pueden estar asociados con la cantidad de patologías musculoesqueléticas reportados por los profesionales de enfermería del HCSL por medio de un enfoque modelado de análisis.

En particular, se busca (1) examinar la asociación

entre el tipo de calzado utilizado y la cantidad de patologías, (2) investigar el efecto de variables demográficas y laborales con la respuesta observada, y (3) determinar si el efecto asociado con el tipo de calzado es modificado por las características demográficas y laborales.

### Materiales y métodos

Para el análisis se utiliza la muestra obtenida por Maidana *et al.* (2020); observada durante el año 2017, por selección aleatoria del conjunto de profesionales de enfermería y obstetricia que prestan servicios en el Hospital Universitario de Clínicas en la ciudad de San Lorenzo, Paraguay. La unidad observacional es el individuo profesional; quién ha dado su consentimiento informado para responder al cuestionario diseñado ad hoc.

El estudio es observacional y transversal. No existe control sobre las variables de interés y los valores observados muestran la situación en un corte temporal específico.

La respuesta de interés es la cantidad de patologías reportadas por cada caso; el factor de interés primario es el tipo de calzado. Se consideran variables demográficas, el sexo, la edad y el estado civil; mientras que son variables laborales, la especialidad (enfermería u obstetricia), la antigüedad y la cantidad de turnos de trabajo.

El análisis se ejecuta utilizando las herramientas disponibles para la especificación, estimación y diagnóstico de modelos estadísticos lineales generalizados (MLGZ).

Los datos en forma de conteos aparecen naturalmente en los estudios de investigación. Un enfoque usual consiste en agrupar según los factores de interés, y aplicar una prueba de asociación, p.e. chi-cuadrado, para extraer conclusiones sobre las asociaciones entre las variables que clasifican los conteos.

Alternativamente, se consideran los conteos observados como respuesta y se construye una ecuación, e.g., un modelo, que describa la relación entre la respuesta y las variables de interés, e.g., las variables explicativas o predictoras.

Una teoría general sobre los MLGZs se puede

encontrar en McCullagh & Nelder, (1998), Agresti, (2002) y Dobson & Barnett, (2008).

En términos generales un MLGZ consta de tres partes; a) la componente aleatoria, b) componente sistemática y c) una función enlace, que relaciona el valor esperado de las observaciones con la componente sistemática o predictor lineal. En la componente aleatoria se define la variable a ser estudiada junto con su distribución de probabilidad. El predictor lineal consiste en la combinación lineal de las variables explicativas cuyos coeficientes se estiman con los datos muestrales, usualmente por medio de una técnica iterativa ponderada basada en la función verosimilitud.

Se procede en dos etapas, 1) Análisis exploratorio, y 2) Búsqueda de una representación adecuada de la relación entre la cantidad de patologías y las variables explicativas de interés. En la primera etapa se utilizan procedimientos gráficos para investigar las relaciones entre variables y las características distribucionales de la respuesta. En la segunda, se parte de una relación entre variables donde la respuesta se considera Poisson estándar; a continuación, se evalúa la presencia de sobredispersión, y finalmente se contrasta la hipótesis que la cantidad de patologías está representada por dos procesos, uno que describe la existencia/ausencia de patologías y otro condicional a que existe patologías, que describe la cantidad de dolencias en cada sujeto; esto es, el proceso para conteos mayores que cero.

Para evaluar la bondad de ajuste se recurre a la deviancia. En realidad, la deviancia mide la discrepancia entre los datos y el modelo, es decir, la falta de ajuste; por lo tanto, un buen modelo debería presentar valores bajos de la deviancia. Para comparar dos modelos tal que uno contiene al otro, e.g., modelos anidados o jerárquicos, se utiliza la reducción de la deviancia, y el estadístico de prueba chi-cuadrado o cociente de verosimilitudes. Cuando los modelos no son jerárquicos se utiliza el criterio de información de Akaike (AIC) o el criterio de información Bayesiano (BIC), ambos equivalentes. El AIC disminuye a medida que el modelo mejora; por tanto, se espera un valor bajo del mismo.

Para la bondad de ajuste y comparación de modelos se utiliza también los llamados rootogramas, originalmente atribuidos a Tukey, (1977). Una amplia discusión de esta herramienta se encuentra en Kleiber & Zeileis (2016).

El modelo Poisson es el estándar para conteos, (Dobson, 2002; Fahrmeir & Tutz, 2001; Fox, 2008; McCullagh & Nelder, 1989; Nelder y Wedderburn, 1972).

Para este modelo, la distribución de la cantidad de patologías es Poisson, cuya característica distintiva es que el valor esperado de los conteos es igual a la varianza. La función que relaciona el valor esperado de los conteos con el predictor lineal es el logaritmo; en otras palabras, el modelo es aditivo en la escala del logaritmo del valor esperado.

Con datos conteos son usuales dos situaciones. En primer lugar, la sobre dispersión o variación extra (Poisson), que se detecta cuando los datos presentan mayor variabilidad que la esperada si las observaciones fueran en realidad Poisson. Segundo, cuando se observan una mayor cantidad de ausencias del evento de interés que la esperada con la suposición Poisson. En ambos casos, el modelo estándar necesita ser corregido.

En el caso de la sobre dispersión, el efecto inmediato es que las varianzas son mayores que las estimadas con el modelo estándar, por lo que las pruebas de hipótesis serán más restrictivas y se encontrarían más variables significativas o mayor significación en las pruebas de hipótesis.

La variación extra se puede originar por ausencia de variables importantes o por falta de independencia de las observaciones o por heterogeneidad en las unidades.

### Modelos con sobredispersión

Se introduce un parámetro, **phi**, adicional para estimar la varianza extra. **Phi** se llama "parámetro de escalamiento de la sobredispersión". El modelo Poisson tiene valor esperado correspondiente al modelo estándar, pero la varianza se multiplica por el parámetro adicional; un valor estimado mayor que uno es señal de sobredispersión. El valor de **phi** se calcula como el cociente del estadístico

chi-cuadrado de Pearson dividido los grados de libertad correspondientes. Con la phi estimada, los intervalos de confianza se amplían en un factor raíz cuadrada de **phi**. (Gardner & Mulvey, 1995; Long, 1997). Como en el modelo estándar, un valor bajo de la deviancia señala un buen modelo.

Un modelo alternativo para datos sobre dispersos es el modelo Binomial negativo. Gardner & Mulvey, 1995; Hilbe, 2014). Este modelo captura el exceso de variación asumiendo que existe una variación no explicada entre los sujetos y con la propiedad de que no tiene efecto sobre el valor promedio de los conteos.

La distribución de probabilidad Binomial negativa además del parámetro promedio tiene adicionalmente un parámetro **theta**. Este parámetro controla el exceso de variación con respecto a la Poisson. A medida que **theta** aumenta, la distribución Binomial negativa tiende a la distribución Poisson; en el límite ambas son equivalentes. Estrictamente, el modelo Binomial negativo no es un MLGZ; pero cuando **theta** es fijo, puede probarse que cumple las condiciones para un modelo generalizado y por tanto se utilizan las técnicas correspondientes.

En general es posible comparar a) el modelo Poisson estándar con el modelo Poisson con variación extra; b) el modelo Poisson estándar con el modelo Binomial Negativo. El modelo sobre disperso Poisson y el modelo BN no pueden compararse directamente pues no están anidados; en este caso, el AIC o BIC son adecuados.

### Exceso de ceros

Un problema frecuente al modelar la cantidad de eventos observados en una unidad es que la cantidad de no-eventos es mayor que la esperada bajo la distribución considerada. En este caso están disponibles dos opciones de uso extendido; **a)** el modelo inflado por ceros (ZIF)-zero-inflated Lambert, (1992), y **b)** modelo hurdle (Hdl) (Cragg, 1971; Mullahy, 1986).

El primer caso es útil en la presencia de ceros estructurales, esto es, cuando existen unidades que no pueden presentar el evento de interés. La segunda opción es aplicable cuando la unidad pudiendo

presentar el evento de interés, en la muestra no lo evidencia.

Estructuralmente ambos tienen dos componentes. Una corresponde a los conteos positivos y otra que modela los ceros. En cada componente se pueden utilizar conjuntos distintos de variables explicativas. En la primera componente son posibles, entre otras, la distribución Poisson y la Binomial negativa; en la segunda una posibilidad es la distribución Binomial para un ajuste del modelo logístico.

## Resultados y discusión

Del proceso de data-QC se mantienen 345 cuestionarios.

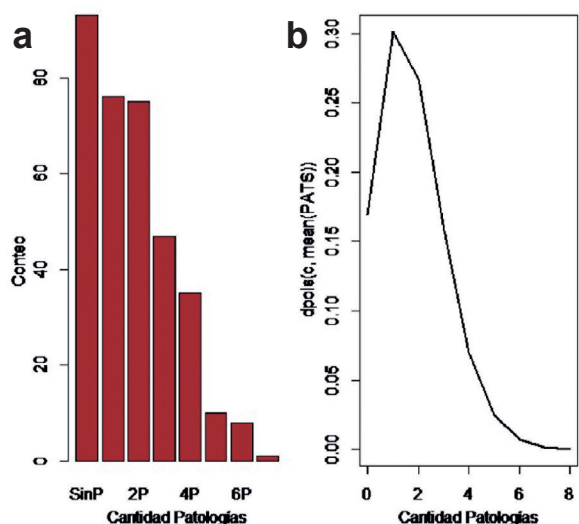
### Variables explicativas

En la muestra han respondido 67 varones, que representan un 20% del total, y 278 mujeres (aprox. 80%). El rango de edad está entre 23 y 59 años, con un promedio de 36.6 años; del total, el cincuenta por ciento tiene menos de 35 años, mientras que el 25% mayor tiene 41 años o más. En el conjunto, 45% está casado y cerca del 50% se reporta soltero; el cinco por ciento restantes se reparte en proporciones casi iguales entre separados y viudos. Con relación a las variables laborales, 326 son profesionales de enfermería y 19 son de obstetricia; del total, un poco más del 58% tiene antigüedad menor a cinco años. La gran mayoría, aproximadamente el 92%, se desempeña en uno (43%) o dos (49%) turnos.

Para el estudio de la variable de interés se consideran cuatro tipos de calzado de uso laboral. El calzado más utilizado es del tipo anatómico con 181 (53%) personas, mientras que el de menor uso es el calzado velcro, que es utilizado por el 11% de los entrevistados; le sigue el de tipo cordón con un 14%, y el tipo crocs es usado para el trabajo por el 22% restante.

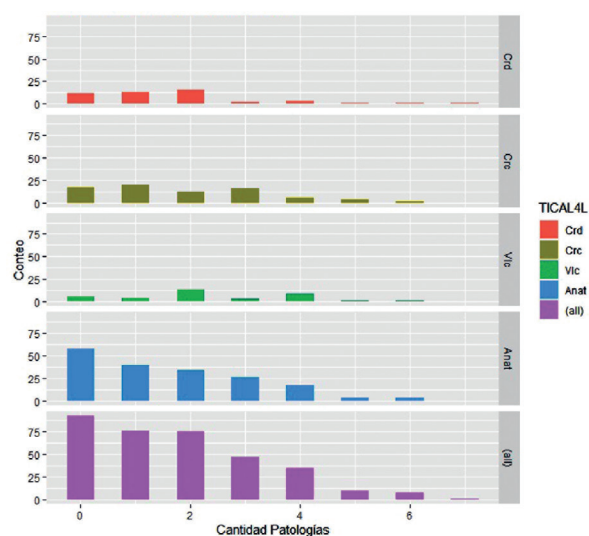
### Cantidad de Patologías

El evento en estudio es la cantidad de patologías musculoesqueléticas declaradas por cada entrevistado. Se considera que los individuos son indepen-



**Figura 1.** Distribución de la cantidad de patologías. a) Distribución de frecuencias. b) Proporción.

dientes entre sí; es decir, la cantidad condiciones en un individuo no hace más o menos improbable la ocurrencia de patologías en otra persona. (Es posible que exista una independencia en bloques, e.g., cátedras distintas son independientes, pero internamente no lo son). Se ha encontrado que aproximadamente un 25% de los 345 entrevistados han manifestado no presentar patología alguna; mientras que el 75%, es decir, tres de cada cuatro, tienen al menos una dolencia del tipo musculoesquelético.



**Figura 2.** Cantidad de patologías por tipo de calzado utilizado.

quelético. La cantidad media de patologías por individuo es 1.774.

En la Fig. 1(a, b) se ve la distribución de los conteos observados; esta puede compararse con la distribución Poisson esperada, Fig. 1(b), con parámetro igual a la tasa media de ocurrencias estimada con la muestra obtenida. Se distingue que la forma es similar, aparte del hecho que en la muestra se han registrado una mayor cantidad de ceros que en la distribución teórica.

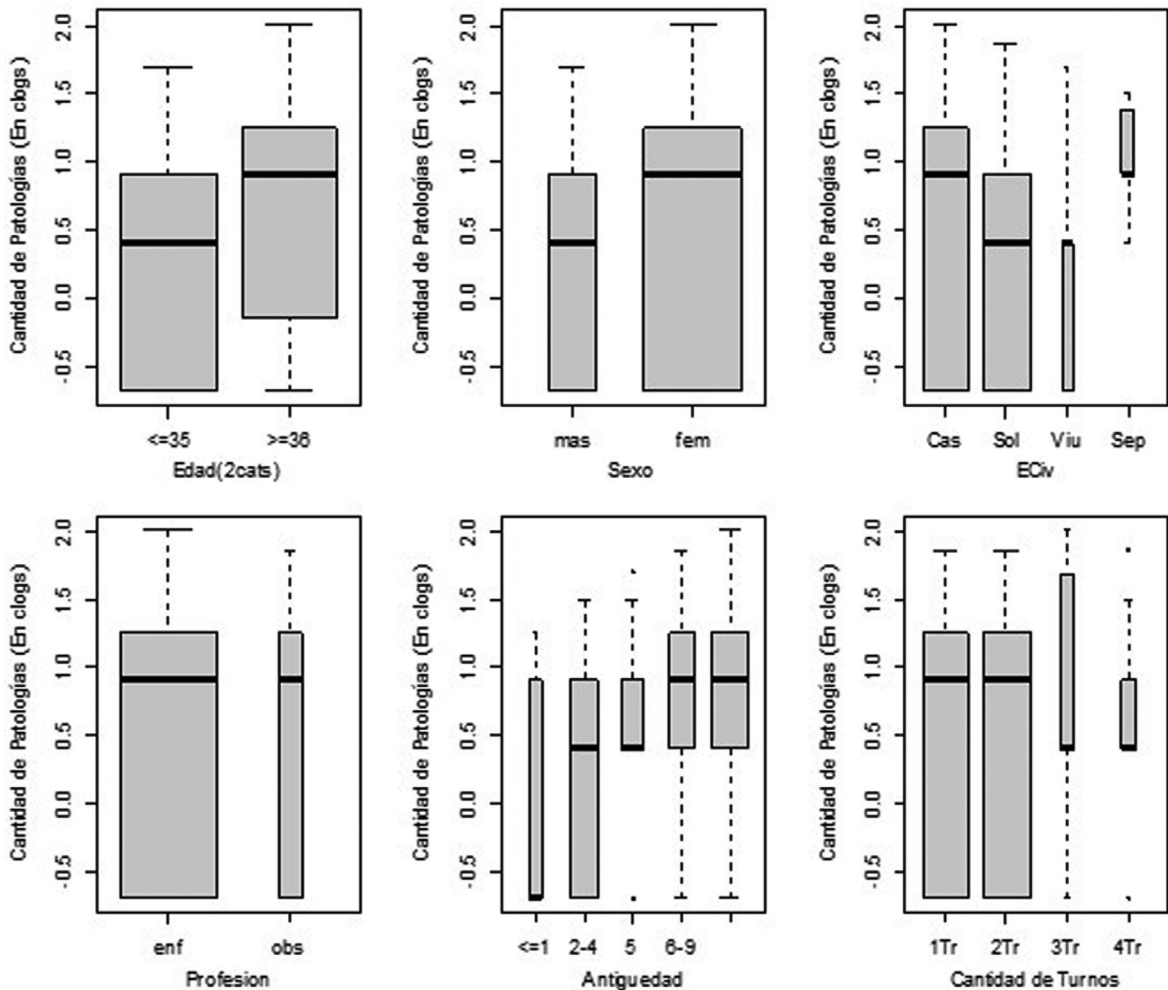
De la distribución de frecuencias de la cantidad de dolencias por individuos, agrupadas por tipo de calzado. En la Fig. 2, se evidencia el desbalance de los datos muestrales; esto es, la cantidad de observaciones por grupo es muy diferente, lo que plantea limitaciones para la estimación de los coeficientes en el modelo y será necesario combinar grupos con objeto de proporcionar mayor estabilidad de las estimaciones.

De la distribución de variables demográficas, edad, sexo y estado civil, se detecta evidencia de que la distribución de los conteos es en cierta medida distinta, ver Fig. 3.

El grupo de 36 años o más tiene la mediana aproximadamente al nivel del tercer cuartil correspondiente al grupo de menor edad; similar estructura se observa en las variables sexo y estado civil. La diferencia observada entre los niveles viudo y separado del grupo Estado civil se establecen con pocas observaciones por lo que no resultan confiables. Del conjunto de variables laborales, profesión y cantidad de turnos tienen estructuras distribucionales similares, por lo que sus respectivos niveles no definen grupos diferentes en términos de cantidad de eventos de interés. En la variable antigüedad, los individuos de menor antigüedad tienen menor cantidad de dolencias que los de cinco años o más; esto sugiere una partición en dos grupos, los de antigüedad menor a cinco y los de cinco o más.

A continuación, se presentan los resultados obtenidos por ajuste de modelos bajo diversas suposiciones. Debe notarse que la muestra es altamente desbalanceada, por lo que existen combinaciones de niveles de los factores que presentan ceros por muestreo; no existen ceros estructurales.





**Figura 3.** Gráfico de cajas y bigotes de la cantidad de patologías clasificadas por: **a)** Edad. **b)** Sexo. **c)** Estado Civil. **d)** Profesión. **e)** Antigüedad. **f)** Cantidad de turnos ejercido.

Del grupo de variables demográficas se toma la variable edad, con dos niveles (menor o igual a 35 años y los mayores a 35); y la variable sexo con dos niveles. Como variable de interés laboral se elige antigüedad tomando en consideración que las otras dos internamente no muestran diferenciación muestral; este factor originalmente con cinco niveles se reagrupa en tres; uno consistente de trabajadores con cinco o menos años de antigüedad, otro con seis a nueve años y el último con diez o más años.

La variable explicativa primaria, originalmente con cuatro niveles, se recodifica para definir dos

grupos; trabajadores que utilizan calzado anatómico y los que no.

Con los niveles redefinidos para cada variable de interés se tienen 24 combinaciones de niveles a ser estudiados.

Una característica típica de los datos conteos es que la varianza tiende a crecer a medida que crece la cantidad de eventos considerados. La suposición básica inicial es que los datos siguen una distribución Poisson; en cuyo caso, el valor esperado de la respuesta es igual a su varianza. Si la varianza es mayor se tiene exceso de variación, técnicamente llamado sobredispersión.

## Relaciones entre cantidad de eventos y variables de interés

Evaluación de las asociaciones entre la cantidad de eventos y las variables explicativas de interés. Los modelos en su mayoría son del tipo lineal generalizado con enlace logaritmo. En el estudio de la variación extra se utiliza la familia cuasi-Poisson. El programa de análisis es R (ver. 4.1.2). Los comandos utilizados y productos seleccionados se pueden encontrar en el Anexo: Modelos Ajustados.

Para la investigación de los predictores, y debido al desbalance estructural, inicialmente se estudian las asociaciones con las variables demográficas incorporando en primer lugar las variables demográficas, sexo, luego edad. A continuación, se compara el posible efecto combinado de ambos. Seguidamente, se incorpora la variable antigüedad y se evalúa si existe efecto combinado de los tres predictores sobre los conteos por individuo.

## Tipo de calzado utilizado (variable: TICAL con dos niveles)

Para evaluar el efecto del tipo de calzado sobre la cantidad de patologías que presenta cada individuo se parte de la suposición de que las dolencias se distribuyen homogéneamente en los individuos.

De la comparación entre el modelo Poisson y el modelo Binomial negativo se concluye que el segundo modelo representa ligeramente mejor a los datos en términos del criterio de información de Akaike, esto es, A-Poisson = 1.234,5 versus A-NB= 1.212,8. Por otra parte, con el modelo BN se reduce ligeramente la significación de la variable TICAL debido a la modificación de la varianza para tomar en cuenta la sobre dispersión; ver Anexo: Modelos Ajustados. Sección: Modelo TICAL. Estos resultados concuerdan con (Maidana *et al.*, 2020).

## Demo y Labor (Poisson estándar)

Utilizando la notación de Wilkinson, (McCullagh & Nelder, 1998) para describir modelos, el modelo completo se escribe como:

$$(M.1) \quad mPoi.DxL = SEXO * EDAD * ANTIG$$

Indicando que se incluye en el modelo, todos los efectos principales, las interacciones de dos variables, por ejemplo, *SEXO:EDAD*, y el efecto combinado de las tres, esto es, *SEXO:EDAD:ANTIG*.

Para el ajuste se utiliza el predictor lineal log (valor esperado de los conteos), y la distribución de error Poisson con enlace logaritmo. Se compara con el modelo aditivo con solo efectos principales,

$$(M.2) \quad mPoi.DL = SEXO + EDAD + ANTIG$$

De la prueba LRT resultante de comparar M.1 con M.2 se determina que las interacciones tienen efectos significativos sobre el logaritmo de los conteos. Como las interacciones de tres variables no superan el corte de significación 5%, estas son excluidas del modelo. Se obtiene entonces el modelo base,

$$(M.3) \quad mPoi.DLi = SEXO+EDAD+ANTIG \\ + SEXO: EDAD + SEXO:ANTIG + \\ EDAD:ANTIG$$

es decir, un modelo con efectos principales y las interacciones de dos variables.

## Demo, Labor y TICAL

- Poisson estándar: Al modelo M.3 se le incorpora el efecto del tipo de calzado para especificar el modelo de investigación. Con esta variable en el modelo, ningún efecto producto de la combinación de tres variables resulta relevante, así como ninguna combinación con el tipo de calzado. Se debe notar que la incorporación de estos nuevos términos al modelo mejora el ajuste significativamente, pero cada uno de forma individual no es relevante.

Por tanto, el modelo Poisson actual se define como,

$$(M.4a) \quad mPoi.DLTi = SEXO + EDAD + \\ ANTIG + TICAL + SEXO:EDAD + \\ SEXO:ANTIG + EDAD:ANTIG$$

Como el modelo Poisson estándar omite la existencia de variación extra, la estimación de la varianza y por tanto los intervalos de confianza

y las pruebas de hipótesis son más exigentes, es necesario evaluar la existencia de sobredispersión.

- b) Chequeo de la variación extra: Por medio de la función "dispersiontest()" del paquete AER y del paquete DHARMA la función testDispersion(), se comprueba que la suposición Poisson de que el parámetro de dispersión es 1, estadísticamente se estima entre 1.17 y 1.26. con un valor-p de 0.008, lo cual evidencia existencia de variación extra-Poisson; ver Fig. 4.

En consecuencia, es conveniente realizar algunas modificaciones para tomar en consideración este hecho.

### Modelo con exceso de variación

Para tomar en cuenta la variación extra, una alternativa es calcular la matriz de varianzas con el parámetro de dispersión estimado; otra posibilidad es ajustar un modelo con familia de distribuciones quasi-Poisson. Alternativamente, se utiliza como distribución del error a la Binomial negativa; para lo cual se recurre a la función glm.nb del paquete MASS. La estructura sistemática es la misma que en (M4a), con la diferencia que ahora la distribución del error es la Binomial negativa. Por tanto,

el modelo NB actual se define como,

$$(M.4b) \quad mBN.DLT_i = SEXO + EDAD + ANTIG + TICAL + SEXO:EDAD + SEXO:ANTIG + EDAD:ANTIG$$

En términos de AIC, los modelos M.4a y M.4b son idénticos; es decir, ambos tienen un ajuste global equivalente; sin embargo, corrigiendo para la variación extra, los parámetros asociados con la edad se duplican, mientras que el coeficiente asociado con el efecto combinado de edad y antigüedad se incrementa en 60%.

### Exceso de ceros

Como se ha visto previamente, la cantidad de trabajadores que no presentan patologías no es consistente con una distribución cuya tasa media de ocurrencias es igual a la observada. Por otra parte, la evidencia de variación adicional podría ser justificada al hecho que el proceso generador de eventos cero es diferente al que genera la ocurrencia de dolencias musculoesqueléticas en los profesionales de enfermería. A los efectos de investigar esta posibilidad, se utiliza un modelo en dos partes; la primera describe la ocurrencia o no del evento de interés, y la segunda modela una distribución de conteos truncada para describir los conteos observados. Aquí se explora el caso en que el mismo conjunto de variables operan en ambas etapas. Para ambos modelos, la generación de cero eventos está gestionada por la distribución binomial, mientras que la componente truncada es descrita ya sea por la distribución Poisson o Binomial negativa. Los resultados se muestran en el Anexo Modelos Ajustados, Sección Exceso de ceros. Los cálculos se realizan utilizando la función "hurdle" del paquete "pscl".

Los dos modelos producen las mismas estimaciones de los coeficientes del modelo; siendo los valores significativos correspondientes en ambos. En el caso Binomial negativo, los valores-p son mayores que su contraparte Poisson. Un aspecto adicional es que el valor del parámetro  $\theta$  es muy grande, por lo que se transforma la varianza de la distribución binomial negativa en el equivalente Poisson; en consecuencia, ambas son iguales.

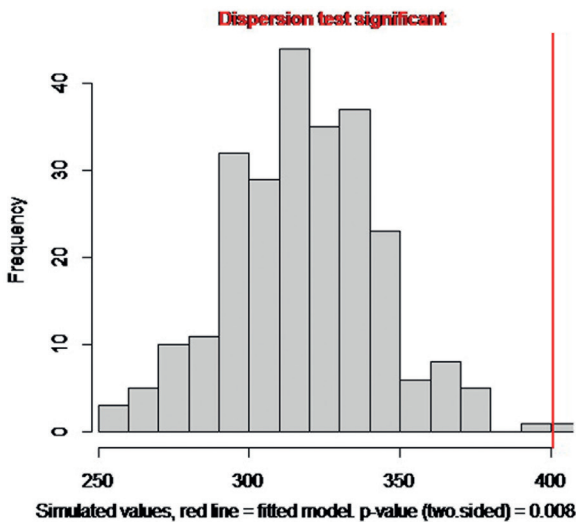


Figura 4. Prueba de sobredispersión, Cantidad de ceros observados es mayor a la cantidad esperada con la suposición Poisson.



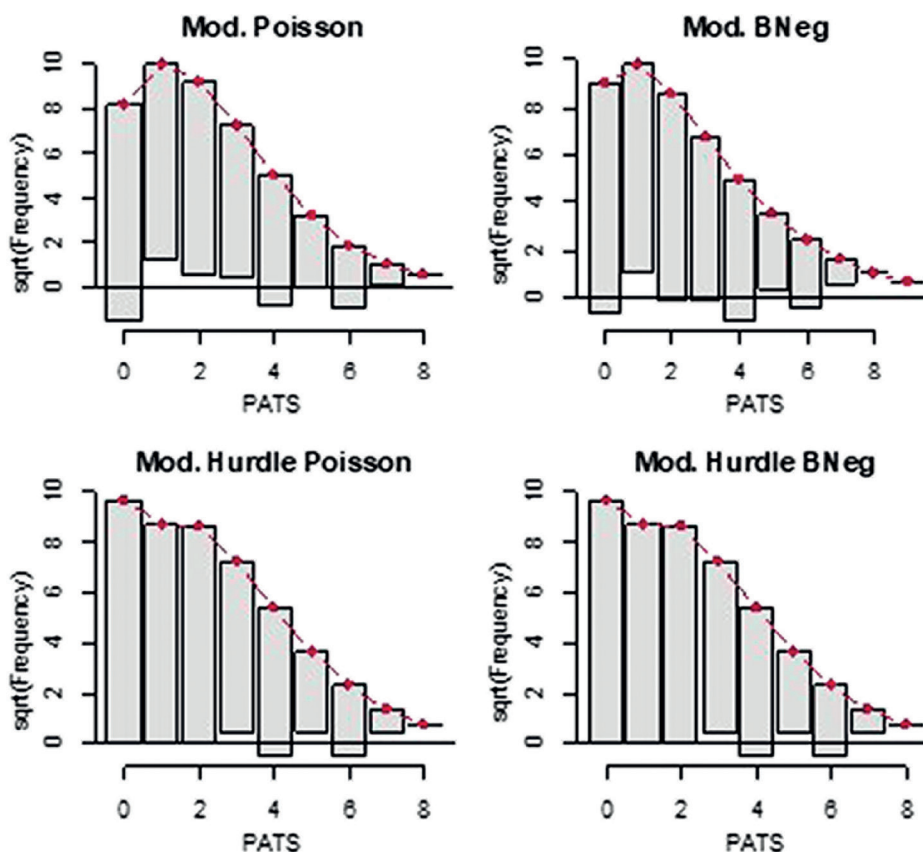


Figura 5. Comparación gráfica de modelos. Rotogramas resultantes para cada modelo ajustado.

### Comparación de modelos

A continuación, se realiza una comparación gráfica, Fig. 5, de los cuatro modelos considerados.

El modelo Poisson presenta un patrón con exceso de ceros, con pocas frecuencias observadas para los valores 1 a 3; los conteos 4 y 6 están sobre ajustados, mientras que los conteos 4 y 7 son reproducidos correctamente

Con la corrección por exceso de variación, el modelo Binomial negativo mejora la representación de los datos; aunque aún muestra los mismos problemas que el modelo Poisson. Los modelos de dos partes mejoran sustancialmente el ajuste. Las estimaciones son consistentes entre ambos modelos. Por la magnitud del parámetro  $\theta$ , resultan equivalentes los modelos "hurdle" Binomial negativo y el modelo "hurdle" Poisson. En consecuencia, se adopta este último para reproducir

el evento en estudio. Es importante mencionar que estos modelos no ajustan adecuadamente (pérdida de ajuste) cuando los conteos son mayores que tres; indicando la necesidad de otras acciones para mejorar el ajuste cuando la cantidad de patologías por persona va aumentando, por ejemplo, incorporar otra variable no considerada, o el ajuste de relaciones cuadráticas, etc.

### Conclusiones

#### a) Efecto TICAL

Del análisis del efecto individual del tipo de calzado utilizado, sin tomar en cuenta la existencia de otros factores, se infiere que esta variable tiene relevancia como factor que incide en la ocurrencia de lesiones musculoesqueléticas. Sin embargo, también se constatan efectos, como la existencia de variación adicional y gran cantidad de individuos sin patolo-

gías, que están asociados con otros factores, y que terminan diluyendo la importancia de esta variable.

#### b) *Efecto Demo, Labor*

Cuando no se considera el efecto debido al tipo de calzado utilizado, La distribución Binomial negativa produce mejor ajuste que la Poisson; en ambos casos existe exceso de ceros no explicados y falta de ajuste.

Los modelos de dos partes (hurdle) mejoran la representación de los datos muestrales; ambos modelos son equivalentes. Se observa pérdida de ajuste cuando la cantidad de eventos por individuo es tres o más. La variable SexFem resulta la única relevante para los conteos (significación 5%). En la parte correspondiente a la mera existencia o no de patologías, resultan significativas la antigüedad y los efectos combinados de sexo y antigüedad.

#### c) *Modifica las variables Demo, Labor el efecto de TICAL*

El modelo considerado adecuado en este trabajo incorpora los factores sexo, edad, antigüedad y tipo de calzado laboral utilizado cotidianamente. La ecuación correspondiente tiene dos partes.

- (1) Modelo para la ocurrencia del evento; Poisson truncado con enlace logaritmo, utilizando solo los coeficientes significativos al 5%.

$$\log(\text{PATS}) = -0.55 + 1.17 \text{ SexFem}$$

- (2) Modelo para conteo 0; Binomial con enlace logit, utilizando solo los coeficientes significativos al 5%

$$\begin{aligned} \text{logit}(\text{PATS}) &= \log(\text{prob}[\text{patologías}] / \text{prob}[\text{no patologías}]) \\ &= -0,55 + 1.17 \text{ SexFem} \\ &+ 2.44 \text{ Antig}(5-9) + 2.76 \text{ Antig}(>=10) \\ &+ (-0.76) \text{ TicalAn} \\ &+ (-1.70) \text{ SexFem:Antig}(5-9) + (-2.32) \\ &\text{SexFem:Antig}(>=10) \end{aligned}$$

#### **Agradecimientos**

A la Lic. Mg. Ada Maidana y su equipo de la Facultad de Enfermería y Obstetricia de la UNA, quienes han diseñado y conducido la toma de muestra y

cuyos datos han sido utilizados en este trabajo.

#### **Referencias**

- Agresti, A. (2003). *Categorical data analysis*. 2<sup>nd</sup> Ed. Hoboken: Wiley. 734pp.
- Anderson, J., Williams, A. & Nester, C. (2016). A narrative review of musculoskeletal problems of the lower extremity and back associated with interface between occupational tasks, feet, footwear, and flooring. *Musculoskeletal Care*, 15(4): 304–315.
- Cragg, J.G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5): 829–844.
- Dobson, A. & Barnett, A. (2008). *Introduction to generalized linear models*. 3<sup>rd</sup> Ed. London: Chapman & Hall / CRC Press. vi + 301pp.
- Fahrmeir, L. & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. 2<sup>nd</sup> Ed. New York: Springer. xxvi + 517 pp.
- Fox, J., (2016). *Applied regression analysis and generalized linear models*. 3<sup>rd</sup> Ed. Thousand Oaks & London: Sage. xxiv + 791 pp.
- Gardner, W. & Mulvey, E. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson and Negative binomial models. *Psychological Bulletin*, 118(3): 392–404.
- Hilbe, J. (2011). *Negative Binomial Regression*. 2<sup>nd</sup> Ed. Cambridge: Cambridge University Press. 550p.
- Kleiber, C., & Zeileis, A. (2016). Visualizing Count Data Regressions Using Rootogramas. *The American Statistician*, 70(3): 296–303.
- Lambert, D. (1992). Zero-Inflated Poisson Regression, with an application to defects in manufacturing. *Technometrics*, 34(1): 1–14.
- Letvak, S., Ruhm, C. (2011). The impact of nurses' health on productivity and quality of care. *The Journal of Nursing Administration*, 41(4): 162–167.
- Long, J. (1997). *Regression models for categorical and limited dependent variables*. Thousand

- Oaks: Sage. *Advanced Quantitative Techniques in the Social Sciences*, 7: xxx + 300 pp.
- Luttmann, A., Jäger, M., Griefahn, B., Caffier, G., Liebers, F. & Steinberg, U. (2003). *Preventing musculoskeletal disorders in the workplace*. Washington, DC: World Health Organization. *Protecting workers' Health*, 5: 32 pp.
- Maidana, A., Mareco, F., Caballero, L., Barrios, E. & Bareiro, J. (2020). Autocuidado en profesional de enfermería con relación al uso de calzado. Hospital de Clínicas, San Lorenzo - Paraguay 2017. *Memorias del Instituto de Investigaciones en Ciencias de la Salud*, 18(3): 24–33.
- McCullagh, P. & Nelder, J. (1989). *Generalized linear models*. 2<sup>nd</sup> Ed. London: Chapman and Hall. 532 pp.
- Mullahy, J. (1986). *Specification and testing of some modified count data models*. *Journal of Econometrics*, 33(3): 341–365.
- Muñoz, C., Rumie, H., Torres, G., & Villarroel, K. (2015). Impacto en la salud mental de la (del) enfermero(a) que otorga cuidados en situaciones estresantes. *Ciencia y Enfermería*, 21(1): 45–53.
- Nelder, J. & Wedderburn, A. (1972). Generalized linear models. *Journal of the Royal Statistical Society (A)*, 135(3): 370–384.
- Reed, L.F., Battistutta, D., Young, J. & Newman, B. (2014). Prevalence and risk factors for foot and ankle musculoskeletal disorders experienced by nurses. *BMC Musculoskeletal Disord*, 15(196): 1–7.
- Tan, C.C. (1991). Occupational health problems among nurses. *Scandinavian Journal of Work, Environment & Health*, 17: 221–30.
- Trinkoff, A.M., Lipscomb, J.A., Geiger-Brown, J. & Brady, B. (2002). Musculoskeletal problems of the neck, shoulder, and back and functional consequences in nurses. *American Journal of Industrial Medicine*, 41(3): 170–178.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley, Reading, MA, USA. xvi + 688 pp.
- Zamora, M., Martínez Alcántara, S., & Balderas López, M. (2019). Trastornos musculoesqueléticos en trabajadores de la manufactura de neumáticos, análisis del proceso de trabajo y riesgo de la actividad. *Acta Universitaria*, 29(e1913): 1–16.

## ANEXO

```

# =====
# ANEXO: MODELOS AJUSTADOS
# -----
# =====

# -----
# -----
# Poisson estandar
# -----
m.0 = glm(PATS ~ 1,
          family = poisson(link= log),
          data = aidaT)
# -----
mPoi.T = glm (PATS ~
              TICAL2L,
              family= poisson(link= log),
              data= aidaT)
summary(mPoi.T)
anova(m.0, mPoi.T, test="Chisq")
# -----
mPoi.T4 = glm (PATS ~
              TICAL4L,
              family= poisson(link= log),
              data= aidaT)
summary(mPoi.T4)
anova(m.0, mPoi.T4, test="Chisq")
# -----
mBN.T = glm.nb (PATS ~
               TICAL2L,
               data= aidaT)
summary(mPoi.T)
anova(m.0, mBN.T, test="Chisq")
# -----
mBN.T4 = glm.nb (PATS ~
                TICAL4L,
                data= aidaT)
summary(mPoi.T4)
anova(m.0, mBN.T4, test="Chisq")

# -----

```

```

# -----
# Binomial Negativa
# -----
library(MASS)
# -----
mBN.0 = glm.nb(PATS ~ 1,
               data = aidaT)
# -----
mBN.T = glm.nb (PATS ~
                TICAL2L,
                data= aidaT)
# -----
summary(mBN.T)
anova(mBN.0, mBN.T, test="Chisq")

# -----
# -----
# Sección: Modelo Demo, Labor (Poisson estándar)
# -----

# (M1)
mPoi.DxL = glm (PATS ~
                SEXOF*EDAD2L*ANTG3F,
                family= poisson(link= log),
                data= aidaT)
summary(mPoi.DxL)
# de este resumen se determina que las interacciones tiene significación menor al
# corte de 5% y por tanto son excluidas del modelo.
# -----

# (M2): Modelo aditivo Demo + Labor
mPoi.DL = glm (PATS ~
               SEXOF + EDAD2L + ANTG3F,
               family= poisson(link= log),
               data= aidaT)
# -----
# Comparación de modelos
anova(mPoi.DxL)
anova(mPoi.DL)
anova(mPoi.DL, mPoi.DxL, test = "LRT")

```



## # Resultados

# -----

	Df	Deviance	Resid.	Df	Resid.	Dev
NULL			344			564.53
SEXOF 1	8.0197		343			556.51
EDAD2L 1	10.1515		342			546.36
ANTG3F 2	9.4514		340			536.91

## Analysis of Deviance Table

Model 1: PATS ~ SEXOF + EDAD2L + ANTG3F

Model 2: PATS ~ SEXOF \* EDAD2L \* ANTG3F

	Resid.	Df	Resid.	Dev	Df	Deviance	Pr(>Chi)
1	340			536.91			
2	333	517.19	7	19.718	0.006212	**	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# -----

## #(M3a) Modelo Demo + Labor, Poisson

```
mPoi.DL3 = glm (PATS ~
  SEXOF + EDAD2L + ANTG3F
  + SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
  family= poisson(link= log),
  data= aidaT)
```

# -----

## #(M3b) Modelo Demo + Labor, NBin

```
mBN.DL3 = glm.nb (PATS ~
  SEXOF + EDAD2L + ANTG3F
  + SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
  data= aidaT)
```

# -----

summary(mPoi.DL3)

summary(mBN.DL3)

# -----

```
mHdlPoi.DL3 <- hurdle(PATS ~
  + SEXOF+EDAD2L+ANTG3F)
```

```

+ SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
data= aidaT,
dist = "poisson",
zero.dist = "binomial")
# -----

mHdIBN.DL3 <- hurdle(PATS ~
+ SEXOF+EDAD2L+ANTG3F+TICAL2L
+ SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
data= aidaT,
dist = "negbin",
zero.dist = "binomial")
summary(mHdIPoi.DL3)
summary(mHdIBN.DL3)

# -----
# Comparación gráfica de modelos
# -----
par(mfrow = c(2, 2))
rootogram(mPoi.DL3, main="Mod. Poisson")
rootogram(mBN.DL3, main= "Mod. BNeg")
rootogram(mHdIPoi.DL3, main="Mod. Hurdle Poisson")
rootogram(mHdIBN.DL3, main= "Mod. Hurdle BNeg")
par(mfrow = c(1, 1))
# -----
# -----

# Seccion: Modelo Demo, Labor, TICAL (Poisson)

# (a) Se incorpora TICAL2L

# (M4a) Modelo Demo, Labor, Tical; 2-var interaccion (Poisson)
mPoi.DLTi = glm (PATS ~
SEXOF + EDAD2L + ANTG3F + TICAL2L
+ SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
family= poisson(link= log),
data= aidaT)
# -----

# Modelo completo, Demo, Labor, Tical (Poisson)

```

```

mPoi.DLTx = glm (PATS ~
  SEXOF*EDAD2L*ANTG3F*TICAL2L,
  family= poisson(link= log),
  data= aidaT)
summary(mPoi.DLTx)

# -----
# Comparación de modelos
# -----
anova(mPoi.DLTi)
anova(mPoi.DLTx)
anova(mPoi.DLTi, mPoi.DLTx, test = "LRT")

# -----
# (b) Chequeo de la variación extra

# Modelo completo Demo, Labor, Tical (Poisson)
m.full= glm (PATS ~
  SEXOF*EDAD2L*ANTG3F*TICAL2L,
  family= poisson(link= log),
  data= aidaT)
# -----

# Testing overdispersion, pkg= AER, or pkg= DHARMA
# -----
library(AER)

# summary(fmp = m.full)
dispersiontest(m.full)

library(DHARMA)
sim.full <- simulateResiduals(m.full, refit=T)
testDispersion(sim.full)

# -----
# -----
# Seccion: Sobredispersión
# -----

# Regresion Quasi-Poisson
# -----

```

```

mqPoi.DLT = glm (PATS ~
                SEXOF+EDAD2L+ANTG3F+TICAL2L
                + SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
                family = quasipoisson(link= "log"),
                data= aidaT)
summary(mqPoi.DLT)

# Regresión Binomial Negativa
# -----

# ajuste con glm.nb de pkg MASS
library(MASS)
# note link= log

# (M4b) Modelo Demo, Labor, Tical; 2-var interaccion (BNeg)
mBN.DLTi <- glm.nb(PATS ~
                  + SEXOF + EDAD2L + ANTG3F + TICAL2L
                  + SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
                  data = aidaT)

summary(mBN.DLTi)

# Resultados
# -----
glm.nb(formula = PATS ~ +SEXOF + EDAD2L + ANTG3F + TICAL2L +
        SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F, data = aidaT,
        init.theta = 5.594638179, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.07654 -1.05573 -0.07816  0.57392  2.31374

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.53446   0.31116  -1.718 0.085863 .
SEXOFfem         0.98515   0.32928   2.992 0.002773 **
EDAD2L>=36       0.15973   0.40546   0.394 0.693619
ANTG3F5-9        1.20338   0.35808   3.361 0.000778 ***
ANTG3F>=10       1.24954   0.40243   3.105 0.001903 **

```

```

TICAL2LAn      -0.25048  0.09515 -2.632 0.008481 **
SEXOfem:EDAD2L>=36  0.35671  0.31641  1.127 0.259580
SEXOfem:ANTG3F5-9 -0.89685  0.38219 -2.347 0.018944 *
SEXOfem:ANTG3F>=10 -1.17230  0.42630 -2.750 0.005961 **
EDAD2L>=36:ANTG3F5-9 -0.63294  0.35081 -1.804 0.071193 .
EDAD2L>=36:ANTG3F>=10 -0.10944  0.36921 -0.296 0.766902

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(5.5946) family taken to be 1)

Null deviance: 451.73 on 344 degrees of freedom

Residual deviance: 410.14 on 334 degrees of freedom

AIC: 1194.6

Number of Fisher Scoring iterations: 1

Theta: 5.59

Std. Err.: 2.01

2 x log-likelihood: -1170.59

```

# -----
# -----
# Seccion: Exceso de ceros
# -----
library(pscl)

mHdlPoi.DLTi <- hurdle(PATS ~
  + SEXOF+EDAD2L+ANTG3F+TICAL2L
  + SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
  data= aidaT,
  dist = "poisson",
  zero.dist = "binomial")
summary(mHdlPoi.DLTi)

# Resultados
# -----

hurdle(formula = PATS ~ +SEXOF + EDAD2L + ANTG3F + TICAL2L +
SEXOF:EDAD2L +
  SEXOF:ANTG3F + EDAD2L:ANTG3F, data = aidaT,
  dist = "poisson", zero.dist = "binomial")

```



Pearson residuals:

Min	1Q	Median	3Q	Max
-1.48153	-0.85848	-0.06537	0.63095	2.87143

Count model coefficients (truncated poisson with log link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.555916	0.566598	-0.981	0.3265
SEXOFfem	1.176632	0.575819	2.043	0.0410 *
EDAD2L>=36	0.273229	0.459408	0.595	0.5520
ANTG3F5-9	1.099569	0.595864	1.845	0.0650 .
ANTG3F>=10	1.034680	0.622511	1.662	0.0965 .
TICAL2LAn	-0.104352	0.097702	-1.068	0.2855
SEXOFfem:EDAD2L>=36	0.199375	0.366102	0.545	0.5860
SEXOFfem:ANTG3F5-9	-0.939229	0.609160	-1.542	0.1231
SEXOFfem:ANTG3F>=10	-1.098317	0.635136	-1.729	0.0838 .
EDAD2L>=36:ANTG3F5-9	-0.336258	0.361610	-0.930	0.3524
EDAD2L>=36:ANTG3F>=10	-0.008171	0.400337	-0.020	0.9837

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.118047	0.468626	-0.252	0.80112
SEXOFfem	1.138580	0.540781	2.105	0.03525 *
EDAD2L>=36	0.009193	0.949543	0.010	0.99228
ANTG3F5-9	2.439441	0.789465	3.090	0.00200 **
ANTG3F>=10	2.761270	1.002130	2.755	0.00586 **
TICAL2LAn	-0.762667	0.269852	-2.826	0.00471 **
SEXOFfem:EDAD2L>=36	0.878560	0.826841	1.063	0.28799
SEXOFfem:ANTG3F5-9	-1.703300	0.844040	-2.018	0.04359 *
SEXOFfem:ANTG3F>=10	-2.323512	1.006838	-2.308	0.02101 *
EDAD2L>=36:ANTG3F5-9	-1.688297	0.950114	-1.777	0.07558 .
EDAD2L>=36:ANTG3F>=10	-0.580604	1.009387	-0.575	0.56515

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 17

Log-likelihood: -573.2 on 22 Df

# -----

mHdIBN.DLTi <- hurdle(PATS ~

```

+ SEXOF+EDAD2L+ANTG3F+TICAL2L
+ SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
data= aidaT,
dist = "negbin",
zero.dist = "binomial")

```

```
summary(mHdIBN.DLTi)
```

```

# Resultados
# -----
# Exceso de Ceros

```

```

hurdle(formula = PATS ~ +SEXOF + EDAD2L + ANTG3F + TICAL2L +
SEXOF:EDAD2L +
SEXOF:ANTG3F + EDAD2L:ANTG3F, data = aidaT,
dist = "poisson", zero.dist = "binomial")

```

Pearson residuals:

```

Min    1Q  Median    3Q    Max
-1.48153 -0.85848 -0.06537  0.63095  2.87143

```

Count model coefficients (truncated poisson with log link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.555916	0.566598	-0.981	0.3265
SEXOFfem	1.176632	0.575819	2.043	0.0410 *
EDAD2L>=36	0.273229	0.459408	0.595	0.5520
ANTG3F5-9	1.099569	0.595864	1.845	0.0650 .
ANTG3F>=10	1.034680	0.622511	1.662	0.0965 .
TICAL2LAn	-0.104352	0.097702	-1.068	0.2855
SEXOFfem:EDAD2L>=36	0.199375	0.366102	0.545	0.5860
SEXOFfem:ANTG3F5-9	-0.939229	0.609160	-1.542	0.1231
SEXOFfem:ANTG3F>=10	-1.098317	0.635136	-1.729	0.0838 .
EDAD2L>=36:ANTG3F5-9	-0.336258	0.361610	-0.930	0.3524
EDAD2L>=36:ANTG3F>=10	-0.008171	0.400337	-0.020	0.9837

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.118047	0.468626	-0.252	0.80112
SEXOFfem	1.138580	0.540781	2.105	0.03525 *
EDAD2L>=36	0.009193	0.949543	0.010	0.99228
ANTG3F5-9	2.439441	0.789465	3.090	0.00200 **
ANTG3F>=10	2.761270	1.002130	2.755	0.00586 **

```

TICAL2LAn      -0.762667  0.269852 -2.826  0.00471 **
SEXOfem:EDAD2L>=36  0.878560  0.826841  1.063  0.28799
SEXOfem:ANTG3F5-9 -1.703300  0.844040 -2.018  0.04359 *
SEXOfem:ANTG3F>=10 -2.323512  1.006838 -2.308  0.02101 *
EDAD2L>=36:ANTG3F5-9 -1.688297  0.950114 -1.777  0.07558 .
EDAD2L>=36:ANTG3F>=10 -0.580604  1.009387 -0.575  0.56515

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Number of iterations in BFGS optimization: 17

Log-likelihood: -573.2 on 22 Df

```

# -----

```

```

mHdIBN.DLT <- hurdle(PATS ~
+ SEXOF+EDAD2L+ANTG3F+TICAL2L
+ SEXOF:EDAD2L + SEXOF:ANTG3F + EDAD2L:ANTG3F,
data= aidaT,
dist = "negbin",
zero.dist = "binomial")

```

```

summary(mHdIBN.DLT)

```

```

# -----

```

```

# Resultados

```

```

# -----

```

```

hurdle(formula = PATS ~ +SEXOF + EDAD2L + ANTG3F + TICAL2L
+ SEXOF:EDAD2L + SEXOF:ANTG3F +
EDAD2L:ANTG3F,

```

```

data = aidaT,
dist = "negbin",
zero.dist = "binomial")

```

Pearson residuals:

```

  Min    1Q  Median    3Q    Max
-1.48153 -0.85848 -0.06536  0.63094  2.87138

```

Count model coefficients (truncated negbin with log link):

```

      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.555832  0.566577 -0.981  0.3266

```

```

SEXOfem          1.176556  0.575800  2.043  0.0410 *
EDAD2L>=36       0.273136  0.459407  0.595  0.5522
ANTG3F5-9        1.099501  0.595845  1.845  0.0650 .
ANTG3F>=10       1.034648  0.622492  1.662  0.0965 .
TICAL2LAn        -0.104349  0.097702 -1.068  0.2855
SEXOfem:EDAD2L>=36 0.199464  0.366103  0.545  0.5859
SEXOfem:ANTG3F5-9 -0.939182  0.609142 -1.542  0.1231
SEXOfem:ANTG3F>=10 -1.098314  0.635121 -1.729  0.0838 .
EDAD2L>=36:ANTG3F5-9 -0.336227  0.361609 -0.930  0.3525
EDAD2L>=36:ANTG3F>=10 -0.008144  0.400338 -0.020  0.9838
Log(theta)       14.681551      NaN    NaN    NaN

```

Zero hurdle model coefficients (binomial with logit link):

```

                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.118047  0.468626 -0.252  0.80112
SEXOfem         1.138580  0.540781  2.105  0.03525 *
EDAD2L>=36      0.009193  0.949543  0.010  0.99228
ANTG3F5-9       2.439441  0.789465  3.090  0.00200 **
ANTG3F>=10      2.761270  1.002130  2.755  0.00586 **
TICAL2LAn       -0.762667  0.269852 -2.826  0.00471 **
SEXOfem:EDAD2L>=36 0.878560  0.826841  1.063  0.28799
SEXOfem:ANTG3F5-9 -1.703300  0.844040 -2.018  0.04359 *
SEXOfem:ANTG3F>=10 -2.323512  1.006838 -2.308  0.02101 *
EDAD2L>=36:ANTG3F5-9 -1.688297  0.950114 -1.777  0.07558 .
EDAD2L>=36:ANTG3F>=10 -0.580604  1.009387 -0.575  0.56515

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Theta: count = 2377477.5223

Number of iterations in BFGS optimization: 52

Log-likelihood: -573.2 on 23 Df

```
# -----
```

```
# -----
```

```
# Comparación de modelos
```

```
# -----
```

```
library(countreg)
```

```
#
```

```
# par(mfrow = c(2, 3))
```

```
par(mfrow = c(2, 2))
rootogram(mPoi.DLTi, main="Mod. Poisson")
rootogram(mBN.DLTi, main= "Mod. BNeg")
rootogram(mHdlPoi.DLTi, main= "Mod. Hurdle Poisson")
rootogram(mHdlBN.DLTi, main= "Mod. Hurdle BNeg")
par(mfrow = c(1, 1))
```

```
# -----
```