



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E
INFORMÁTICA
ESTADÍSTICA

**EVALUACIÓN DE ALTERNATIVAS DE
ANÁLISIS EN EXPERIMENTOS
AGRÍCOLAS BAJO CONDICIONES NO
NORMALES Y MUESTRAS PEQUEÑAS**

ARWELL NATHÁN LEYVA CHÁVEZ

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, ESTADO DE MÉXICO
2021

La presente tesis titulada: **Evaluación de alternativas de análisis en experimentos agrícolas bajo condiciones no normales y muestras pequeñas**, realizada por el alumno: **Arwell Nathán Leyva Chávez**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO



Dr. Gustavo Ramírez Valverde

ASESOR



Dr. Paulino Pérez Rodríguez

ASESOR



Dr. Benito Ramírez Valverde

Montecillo, Texcoco, Estado de México, enero de 2021.

EVALUACIÓN DE ALTERNATIVAS DE ANÁLISIS EN EXPERIMENTOS AGRÍCOLAS BAJO CONDICIONES NO NORMALES Y MUESTRAS PEQUEÑAS

Arwell Nathán Leyva Chávez, M.C.

Colegio de Postgraduados, 2021

RESUMEN

En experimentos agrícolas es común encontrar experimentos donde la variable respuesta es de conteos, éstos se han analizado tradicionalmente con una prueba F (ANOVA) y el Modelo Lineal General (LM), sin embargo, muchas veces no se cumplen los supuestos de normalidad y homogeneidad de varianzas. Una alternativa de análisis es el uso de transformaciones, donde para el caso de variables de conteos, se recomienda la transformación raíz cuadrada, otra es mediante Modelo Lineal Generalizado (GLM) considerando que la variable de conteos tiene distribución Poisson, sin embargo, las distribuciones de los parámetros estimados que se utilizan para hacer inferencia se basan en teoría asintótica y los experimentos utilizan frecuentemente un número pequeño de repeticiones, sobre todo en áreas como la agronomía. Además, cuando se utiliza la distribución Poisson se establece que la media es igual a la varianza, sin embargo, en muchos casos la variable de conteos presenta varianza muy diferente a la media; éste es el caso de sobredispersión; dos de los principales enfoques para el análisis de variables de conteos en presencia de sobredispersión son, utilizar la distribución Binomial negativa o Quasipoisson. En este estudio se comparó el tamaño de la prueba y la potencia para probar el efecto de los tratamientos bajo el enfoque del LM (prueba de F en los modelos Normal y con la transformación raíz cuadrada), y GLM (regresión Poisson, Quasipoisson y Binomial negativa) cuando se tiene variable respuesta de conteos y muestras pequeñas. Mediante simulación Monte Carlo se estudió el efecto del número de repeticiones, tamaño del efecto y el número de tratamientos en la potencia y el tamaño de la prueba de los modelos mencionados sin, y con sobredispersión. Cuando los datos no presentan sobredispersión se encontró que todos los enfoques mantuvieron el valor nominal de su tamaño de prueba, excepto Quasipoisson, el cual tuvo valores mayores a los nominales con una muestra pequeña, sólo para tamaño de muestra cincuenta se mantuvieron cercanos a los nominales; En cuanto a la potencia de la prueba, el modelo Poisson presentó la mayor. En presencia de sobredispersión, el modelo Poisson no mantuvo el tamaño de prueba cercano al nominal, los modelos Binomial negativo y Quasipoisson solamente en muestras grandes (cincuenta repeticiones); en este caso los modelos Normal y Normal transformado mantuvieron el tamaño de la prueba cercano al nominal, sin embargo, en muestras grandes tuvieron menos potencia que el modelo Binomial negativo. La potencia de la prueba al momento de tener una sobredispersión para muestras pequeñas es muy baja, solamente para cincuenta repeticiones

se puede observar que detecta la sobredispersión. Debido a los resultados obtenidos de la prueba de sobredispersión, y que el modelo Poisson y Normal de manera general tuvieron el mismo comportamiento, se recomienda utilizar el modelo Normal para todos los casos presentados.

Palabras clave: Muestras pequeñas, no normales, agronomía, comparación de modelos.

EVALUATION OF ANALYSIS ALTERNATIVES IN AGRICULTURAL EXPERIMENTS UNDER NO NORMAL CONDITIONS AND SMALL SAMPLES

Arwell Nathán Leyva Chávez, M.C.

Colegio de Postgraduados, 2021

ABSTRACT

In agricultural experiments, it is common to find research works where the response variable is counts, these experiments have traditionally been analyzed with a test of F (ANOVA) and the General Linear Model (LM), however, many times they are not fulfilled the assumptions of normality and homogeneity of variances. One of the analysis alternatives is the use of transformations, where for the case of count variables, the square root transformation is recommended, another of the forms is through the Generalized Linear Model (GLM) considering that the count variable has a Poisson distribution, however, this model is based on asymptotic theory and experiments frequently use a small number of replications, especially in areas such as agronomy. In addition to this, when the Poisson distribution is used, the characteristic is established that the mean is equal to the variance, however, in many cases the variable of counts presents variance that is very different from the mean; this is the case of overdispersion (the variance greater than the mean); Two of the main approaches for the analysis of count variables in the presence of overdispersion, are to use the negative Binomial or Quasipoisson distribution. In this study, the test size and power were compared to test the effect of the treatments under the LM approach (F test in Normal models, and F test with the square root transformation), and GLM (Poisson, Quasipoisson and negative Binomial regression) when there is a response variable of counts and small samples. Using Monte Carlo simulation, the effect of the number of repetitions, effect size and the number of treatments on the power and size of the test of the models mentioned above without, and with overdispersion, was studied. When the data do not present overdispersion, it was found that all the approaches maintained the nominal value of their test size, except Quasipoisson, which had test sizes greater than the nominal ones when there was a small sample, only for sample size fifty, kept test sizes close to nominal; Regarding the power of the test, the Poisson model presented the highest. In the presence of overdispersion, the Poisson model did not keep the test size close to the nominal, the negative Binomial and Quasipoisson models only kept it in large samples (fifty replicates); In this case, the Normal and Transformed Normal models kept the test size close to the nominal, however, in large samples they had less power than the negative Binomial model.

Keywords: Small samples, agronomy, model comparison.

DEDICATORIA

A mis hijos y esposa.

AGRADECIMIENTOS

A la Facultad de Ciencias Agrotecnológicas (FACIATEC) de la Universidad Autónoma de Chihuahua (UACH) por el apoyo brindado para la realización de estos estudios.

Al posgrado PSEI-ESTADÍSTICA del Colegio de Postgraduados por permitirme continuar con mi formación académica.

A mi consejo particular por el tiempo invertido en la revisión de ésta investigación.

A cada uno de mis profesores.

A todo el personal administrativo que labora en pro del posgrado, sobretodo por su amabilidad y calidez humana.

A toda mi generación por la ayuda y el apoyo brindado a lo largo de estos años.

CONTENIDO

RESUMEN	iv
ABSTRACT	v
DEDICATORIA	vi
AGRADECIMIENTOS	vii
LISTA DE FIGURAS	x
LISTA DE CUADROS	xiii
1. INTRODUCCIÓN	1
1.1. Planteamiento del problema	3
1.2. Objetivos	3
1.2.1. Objetivo general	3
1.2.2. Objetivos específicos	3
1.3. Hipótesis	4
2. REVISIÓN DE LITERATURA	5
2.1. Diseños experimentales en agronomía	5
2.2. Conceptos estadísticos importantes	6
2.2.1. Distribuciones	6
2.2.2. Potencia de una prueba	7
2.2.3. Tamaño de la prueba	10
2.2.4. Nivel de significancia	10
2.2.5. Tamaño del efecto	11
2.3. Modelo lineal general (LM)	12
2.3.1. Supuestos	13
2.3.2. Modelo y Análisis de Varianza (ANOVA)	15
2.4. Alternativas de análisis	19

2.4.1.	Transformaciones	19
2.4.2.	Modelos lineales generalizados (GLM)	23
2.4.3.	GLM Poisson	32
2.4.4.	GLM Binomial negativo	36
2.5.	Prueba de sobredispersión	40
3.	MATERIALES Y MÉTODOS	41
3.1.	Factores de estudio	42
3.1.1.	Modelos	42
3.1.2.	Número de repeticiones	42
3.1.3.	Número de tratamientos	42
3.1.4.	Tamaño del efecto	43
3.1.5.	Distribución entre medias	43
3.1.6.	Sobredispersión	43
4.	RESULTADOS Y DISCUSIÓN	45
4.1.	Tamaño de la prueba	45
4.1.1.	Sin sobredispersión (datos Poisson)	45
4.1.2.	Sobredispersión grande	48
4.1.3.	Sobredispersión mediana y pequeña	51
4.2.	Potencia	54
4.2.1.	Sin sobredispersión	54
4.2.2.	Sobredispersión grande	64
4.2.3.	Sobredispersión mediana	69
4.2.4.	Sobredispersión pequeña	72
4.3.	Prueba de sobredispersión	74
5.	CONCLUSIONES	76
6.	LITERATURA CITADA	77
	ANEXOS	82

LISTA DE FIGURAS

4.1. Comparación del tamaño de la prueba promedio de los modelos para datos Poisson a través de diferentes repeticiones.	46
4.2. Comparación del tamaño de la prueba promedio de los tratamientos para datos Poisson a través de diferentes repeticiones.	47
4.3. Comparación del tamaño de la prueba promedio de las distintas medias para datos Poisson a través de diferentes repeticiones.	48
4.4. Comparación del tamaño de la prueba promedio de los modelos para datos con sobredispersión grande a través de diferentes tratamientos.	49
4.5. Comparación del tamaño de la prueba promedio de los modelos para datos con sobredispersión grande a través de diferentes repeticiones.	50
4.6. Comparación del tamaño de la prueba promedio de los tratamientos para datos con sobredispersión grande a través de diferentes repeticiones.	51
4.7. Comparación del tamaño de la prueba promedio de los modelos a través de la sobredispersión, caso de ocho tratamientos.	52
4.8. Comparación del tamaño de la prueba promedio de los modelos a través de la sobredispersión, caso de cuatro repeticiones.	53
4.9. Comparación del tamaño de la prueba promedio de los diferentes tratamientos a través de la sobredispersión, caso de cuatro repeticiones.	54
4.10. Comparación de la potencia de la prueba de modelos para datos Poisson a través del tamaño del efecto.	56
4.11. Comparación de la potencia de la prueba promedio de modelos para datos Poisson a través de los tratamientos, caso del tamaño de efecto grande.	57
4.12. Comparación de la potencia de la prueba promedio de las diferentes repeticiones para datos Poisson a través del tamaño del efecto.	58
4.13. Comparación de la potencia de la prueba de los tratamientos para datos Poisson a través de las diferentes repeticiones, caso de medias balanceadas y tamaño del efecto mediano.	58

4.14. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de las diferentes repeticiones.	59
4.15. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de los tratamientos, caso de cuatro repeticiones.	60
4.16. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través de las diferentes repeticiones, caso de cuatro tratamientos.	61
4.17. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través de las diferentes repeticiones, caso de ocho tratamientos.	62
4.18. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través del tamaño del efecto, caso de cuatro tratamientos.	63
4.19. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de las diferentes medias, caso de ocho tratamientos.	64
4.20. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través del tamaño del efecto.	65
4.21. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de los tratamientos, caso de medias balanceadas y tamaño del efecto mediano.	66
4.22. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de las repeticiones.	67
4.23. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de los tratamientos, caso de ocho repeticiones y tamaño de efecto pequeño.	67
4.24. Comparación de la potencia de la prueba de las repeticiones para datos con sobredispersión grande a través de los tamaños del efecto.	68
4.25. Comparación de la potencia de la prueba de los tratamientos para datos con sobredispersión grande a través de las repeticiones, caso de medias balanceadas y tamaño del efecto mediano.	69
4.26. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión mediana a través de los tamaño del efecto.	70
4.27. Comparación de la potencia de la prueba de las diferentes repeticiones para datos con sobredispersión mediana a través de los tamaño del efecto.	70
4.28. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión mediana a través de los tratamientos, caso de tamaño del efecto mediano y medias balanceadas.	71

4.29. Comparación de la potencia de la prueba de los tratamientos para datos con sobredispersión mediana a través de las repeticiones, caso de tamaño del efecto mediano y medias balanceadas.	72
4.30. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión pequeña a través de las repeticiones.	73
4.31. Comparación de la potencia de la prueba de las repeticiones para datos con sobredispersión pequeña a través del tamaño del efecto.	74
4.32. Comparación de la potencia de la prueba de sobredispersión para datos Poisson a través de las repeticiones, caso de medias balanceadas, cuatro tratamientos y tamaño de efecto pequeño.	74
4.33. Comparación de la potencia de la prueba de sobredispersión para datos Poisson a través de los tratamientos, caso de medias balanceadas, cuatro repeticiones y tamaño de efecto pequeño.	75

LISTA DE CUADROS

2.1. Resultados de las pruebas estadísticas.	8
2.2. Algunas convenciones para definir al tamaño del efecto.	11
2.3. Algunas transformaciones conocidas presentadas por Hinkelmann y Kemphorne (1994).	22
3.1. Parámetros de forma de la sobredispersión (k) utilizados.	43
5.1. Valores de las medias para datos Poisson.	82
5.2. Valores de las medias para datos con sobredispersión grande.	83
5.3. Valores de las medias para datos con sobredispersión mediana.	83
5.4. Valores de las medias para datos con sobredispersión pequeña.	84
5.5. Datos utilizados en la investigación.	84
5.5. Datos utilizados en la investigación.	85
5.5. Datos utilizados en la investigación.	86
5.5. Datos utilizados en la investigación.	87
5.5. Datos utilizados en la investigación.	88
5.6. Valores de potencia de los modelos utilizados en la investigación.	88
5.6. Valores de potencia de los modelos utilizados en la investigación.	89
5.6. Valores de potencia de los modelos utilizados en la investigación.	90
5.6. Valores de potencia de los modelos utilizados en la investigación.	91
5.6. Valores de potencia de los modelos utilizados en la investigación.	92

1. INTRODUCCIÓN

Una de las necesidades básicas de los seres humanos, como lo es comer, está directamente ligada con el campo, el creciente desarrollo poblacional nos debe de preocupar y ocupar. En los últimos años, se han tenido avances en las metodologías estadísticas, esto, ligado a los avances tecnológicos, los cuales, impactan en las investigaciones agronómicas y de recursos naturales limitados con los que contamos como sociedad. Esto se ve en análisis cada vez más robustos, con datos con variables de respuesta categóricas o continuas, pero que no necesariamente impliquen normalidad, datos tomados a lo largo de cierto tiempo, espaciales, multi ubicación o multi año.

Los datos recabados pueden estar directamente relacionados con el objetivo del experimento o pueden ser datos complementarios que pueden ayudar en la interpretación de los resultados que realmente nos interesan. Es necesaria la evaluación asertiva de cualquier experimento para contar con resultados claros y aplicables, de aquí la necesidad de los métodos estadísticos relacionados con el cultivo de la tierra. El objetivo principal de los experimentos agrícolas es determinar si una variable en particular tiene un efecto sobre otra variable respuesta, en ocasiones, en ésta área del conocimiento se trabaja con variables de respuesta de conteos.

Frecuentemente en experimentos agrícolas, los datos de conteos se analizan mediante una prueba F de un ANOVA basado en un LM, el cual es una forma alternativa para probar hipótesis que se basa en la partición de la variabilidad, para estudiar relaciones entre dos o más variables en las ciencias experimentales.

Los supuestos que se deben de cumplir para utilizar el ANOVA son la independencia, asignando de manera aleatoria los tratamientos a las unidades experimentales (UE), homogeneidad de varianzas, donde lo deseable es que la variación dentro de cada tratamiento sea similar, y que las variables deben tener una distribución Normal. Éstas premisas del ANOVA [Bandera-Fernández y Pérez-Pelea \(2018\)](#) nos comentan que son supuestos, con el fundamento de que si el número de observaciones es elevado, los datos pueden llegar a ajustarse a

la normalidad, en la realidad esto es complicado debido a que los recursos empleados en agronomía son costosos. Además, [Shadish y Sullivan \(2011\)](#) han informado que las variables de respuesta medidas en diseños experimentales son muy a menudo de naturaleza discreta en vez de continua, y para estos resultados discretos no se cumple el supuesto de normalidad condicional.

Las transformaciones es una de las alternativas utilizadas al momento de que fallan los supuestos del LM, donde la más utilizada para datos de conteos, es la transformación raíz cuadrada, la cual, es un caso particular de la transformación de Box-Cox, para poder cumplir con el supuesto de homogeneidad de varianzas, por lo que, permite obtener análisis aproximados utilizando los métodos estándar de la teoría normal, con el costo en la interpretabilidad.

Por otra parte, [Wang y col. \(2015\)](#) comentan que los datos que son colectados en las investigaciones agrícolas no satisfacen las premisas de los LM, por lo que una alternativa son los Modelos Lineales Generalizados (GLM por sus siglas en inglés) que proporcionan una vía de análisis que no requiere el supuesto de normalidad. Los métodos estadísticos frecuentemente utilizados se han desarrollado principalmente para datos con distribución Normal. Los GLM amplían la teoría para incluir una extensa clase de distribuciones, incluyendo aquellas que comúnmente se utilizan para conteos, proporciones y distribuciones sesgadas ([Gbur y col., 2012](#)).

Los GLM proporcionan las herramientas para realizar análisis de los datos colectados en los diseños experimentales, adaptados de manera particular a las necesidades de los estudios, los cuales, son diferentes a los diseños experimentales convencionales, esto debido a que dan una metodología que permite el análisis de problemas con observaciones que tienen una estructura distribucional perteneciente a la familia exponencial abarcando una gran cantidad de situaciones ([Nelder y Wedderburn, 1972](#)). De esta manera, los GLM nos ayudan a replantear la práctica estadística en la investigación agronómica y de recursos naturales unificando una amplia variedad de métodos estadísticos.

En una distribución Poisson, se tiene la característica de que la media y la varianza son iguales, pero existen ocasiones cuando trabajamos con conteos que no suele ser cierta esta hipótesis, en el caso donde la varianza es mayor a la media, se dice que se tiene sobredispersión y el GLM Poisson no se debe utilizar, para esto, existen los modelos Quasipoisson o Binomial negativo como alternativas de análisis.

1.1. Planteamiento del problema

A menudo, los estudios de diseños experimentales consideran resultados de naturaleza discreta en lugar de continua, como recuentos, porcentajes o tasas, en esos casos, el supuesto de normalidad no se cumple. [Qumsiyeh y Shaughnessy \(2012\)](#) comenta que una situación común ocurre cuando las respuestas son de naturaleza discreta, por ejemplo, conteos. Una forma de analizar estos datos experimentales es utilizar una transformación para las respuestas; otra es utilizar una función de enlace basada en un enfoque de GLM.

Debido a que existen diferentes formas de abordar una variable respuesta de conteos en experimentos agrícolas, se analizarán los enfoques alternativos de modelos LM, utilizados más frecuentemente en investigación, y los GLM, ya que ignorar los supuestos de los modelos de las metodologías estadísticas, puede ocasionar que las inferencias sean incorrectas.

1.2. Objetivos

En este apartado se presentan los objetivos general y particulares que se abarcan en esta investigación.

1.2.1. Objetivo general

Comparar el tamaño de la prueba y la potencia de análisis alternativos para estudiar experimentos con variables de conteos como variable respuesta.

1.2.2. Objetivos específicos

1. Evaluar y comparar el tamaño de las pruebas de distintos enfoques de análisis alternativos en experimentos agrícolas con variable respuesta de conteos y muestras pequeñas, sin, y con sobredispersión.
2. Evaluar y comparar la potencia de la prueba de distintos enfoques de análisis alternativos en experimentos agrícolas con variable respuesta de conteos y muestras pequeñas, sin, y con sobredispersión.
3. Evaluar y comparar la potencia y el tamaño de las pruebas de los distintos enfoques de análisis alternativos de experimentos agrícolas con variable respuesta de conteos y muestras pequeñas, sin, y con sobredispersión, con diferentes repeticiones.

4. Evaluar y comparar la potencia y el tamaño de las pruebas de los distintos enfoques de análisis alternativos de experimentos agrícolas con variable respuesta de conteos y muestras pequeñas, sin, y con sobredispersión, con diferentes tratamientos.
5. Evaluar y comparar la potencia y el tamaño de las pruebas de los distintos enfoques de análisis alternativos de experimentos agrícolas con variable respuesta de conteos y muestras pequeñas, sin, y con sobredispersión, con diferentes tamaños de efecto.

1.3. Hipótesis

Los GLM Poisson, Quasipoisson y Binomial negativo tienen mejor comportamiento en cuanto a tamaño de las pruebas y potencia que los LM Normal y Normal transformado al analizar variables de conteos con, y sin sobredispersión.

2. REVISIÓN DE LITERATURA

2.1. Diseños experimentales en agronomía

La investigación relacionada con el campo agrícola, pecuario y veterinario es una necesidad apremiante en el país y el mundo ([Gabriel y col., 2017](#)), los diseños experimentales son utilizados en diferentes áreas del conocimiento, como lo son la industria, la mercadotecnia, la ecología, la medicina, la agronomía, por mencionar algunos; la fase de los diseños es esencial en el desarrollo de un estudio experimental, por lo que constituyen la principal herramienta en la investigación agrícola. Debemos de tomar en cuenta que, en el área agronómica los fenómenos biológicos pueden aportar una gran variabilidad al experimento, es una característica muy importante y que no debemos dejar de lado, puesto que afecta en gran medida al resultado, debido a que los factores incontrollables que afectan a estos fenómenos, son muchos e impredecibles.

Es fundamental en un experimento establecer las variables a analizar, los factores deben ser definidos en función de los resultados que se desean alcanzar, entre más factores se deseen analizar, mayores recursos necesitaremos, en su gran mayoría, un experimento agrícola conlleva tiempo y dinero, por lo que el material experimental debe ser cuidadosamente seleccionado a fin de garantizar la calidad en los resultados, tomando en cuenta que, la potencia estadística esta ligada al número de repeticiones, por lo que un experimento con buena potencia requerirá un buen tamaño de muestra.

Un evento de conteos se refiere al número de veces que ocurre un evento, es decir, es la realización de una variable aleatoria de valor entero no negativo, en agronomía por ejemplo, podemos encontrar muchas variables en este sentido, como lo son el número de frutos, número de insectos, número de hojas, etc.

[Kempton y col. \(1996\)](#) manifiesta que, dentro de los conceptos importantes a considerar para un buen diseño de experimentos se encuentra la aleatorización, la variabilidad y la correcta elección de niveles para los factores a considerar en el estudio y que factores

adicionales, que representan una variación controlada en el entorno en el que se realizará el experimento, pueden incluirse ocasionalmente en los ensayos.

Para poder mejorar los análisis, se puede hacer uso de información adicional, con la cual a menudo se cuenta y nos ayudará a reducir el error en el experimento, a éstas variables se les llama covariables, ya que no son factores que sean de interés estudiar, pero se deben de tomar en cuenta para mejorar la precisión de las diferencias que pudieran existir entre tratamientos, esto se lleva a cabo a través de un análisis de covarianza (ANCOVA).

2.2. Conceptos estadísticos importantes

2.2.1. Distribuciones

2.2.1.1. Distribución Binomial

Se dice que una variable aleatoria discreta X tiene distribución binomial si su función de densidad es tal que

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{de otro modo} \end{cases}, \quad (2.1)$$

cuando el parámetro p satisface $0 \leq p \leq 1$. $1 - p$ es usualmente denotado por q .

2.2.1.2. Distribución Binomial Negativa

Se dice que una variable aleatoria discreta X tiene distribución binomial negativa si su función de densidad esta dada por

$$f_X(x) = \begin{cases} \binom{x+r-1}{r-1} (1-p)^r p^{x-1}, & x = 0, 1, 2, \dots \\ 0, & \text{de otro modo} \end{cases}, \quad (2.2)$$

donde $r \in \mathbb{N}$ y $p \in [0, 1]$.

2.2.1.3. Distribución Poisson

La variable aleatoria que utiliza la distribución Poisson es la discreta, la cual toma valores enteros positivos. Se dice que una variable aleatoria discreta X tiene distribución Poisson si

su función de densidad es tal que

$$f_X(x) = \begin{cases} \frac{\exp^{-\lambda} \lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{de otro modo} \end{cases}, \quad (2.3)$$

donde $\lambda > 0$.

2.2.1.4. Distribución Normal

Se dice que una variable aleatoria continua X tiene distribución normal con parámetros μ y σ^2 si la función de densidad de X es de la forma

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, x \in \mathbb{R}, \quad (2.4)$$

donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

2.2.1.5. Distribución Gamma

Se dice que una variable aleatoria continua X tiene distribución gamma si su función de densidad es tal que

$$f_X(x) = \frac{x^{\alpha-1} \exp^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}, x > 0, \quad (2.5)$$

donde $\alpha > 0, \beta > 0$.

2.2.2. Potencia de una prueba

Una prueba de hipótesis es un procedimiento estadístico matemático que nos permite probar una aseveración acerca del parámetro de una población. Una hipótesis estadística es una afirmación o conjetura sobre la distribución de una o más variables aleatorias. Si la hipótesis estadística especifica completamente la distribución, entonces se llama simple; de lo contrario, es compuesta. Una prueba de hipótesis estadística (\mathcal{H}) es una regla o procedimiento para decidir si rechazar \mathcal{H} (Mood y col., 1974). De aquí en adelante se utilizará la letra capital Υ para denotar una prueba.

En muchos problemas de pruebas de hipótesis se discuten dos, la primera, es la hipótesis que se está probando, que se llama la hipótesis nula denotada por \mathcal{H}_0 y la segunda que se llama hipótesis alternativa, denotada por \mathcal{H}_a . La idea general es que si la hipótesis nula es

falsa, entonces la alternativa es verdadera y viceversa, debido a esto, pueden existir dos tipos de errores.

El interés es poder comprender y minimizar los errores en la inferencia estadística, como vemos en la tabla 2.1. El primer error es pensar que los resultados de la muestra no tienen un efecto real en la población, cuando en el experimento, si tienen algún tipo de efecto, a esto se le conoce como un error tipo I. Por ejemplo, suponga que se tiene un experimento en el cual se aplican cuatro tratamientos a una población y se desea probar si existe efecto de tratamiento. Para ello se seleccionan n individuos en la población en forma aleatoria. Posteriormente a n_1 de ellos se les aplica el tratamiento uno, a n_2 el tratamiento 2, a n_3 el tratamiento 3 y al resto se les da un placebo, y se anota el rendimiento al final de la temporada; el error tipo I es decir que alguno de los tratamientos aplicados mejora el rendimiento, cuando en realidad no lo hace.

Tabla 2.1. Resultados de las pruebas estadísticas.

Conclusión a la que se llegó en		¿Qué es verdad en la Población?	
		El tratamiento no tiene efecto	El tratamiento tiene efecto
el estudio	Sin efecto	Conclusión correcta ($p = 1 - \alpha$)	Error tipo II ($p = \beta$)
	Efecto de tratamiento	Error tipo I ($p = \alpha$)	Conclusión correcta ($p = 1 - \beta$) Potencia

Un segundo tipo de error, denominado error tipo II, también es común en las pruebas de hipótesis estadísticas (Sedlmeier y Gigerenzer, 1992; Cohen, 1994). Un error tipo II ocurre cuando los investigadores concluyen a favor de \mathcal{H}_0 , cuando de hecho \mathcal{H}_a es cierta, para nuestro ejemplo, es no recomendar algún tratamiento cuando en realidad si existe diferencias en los rendimientos. El análisis estadístico de potencia se refiere a los errores tipo II. La potencia de una prueba estadística se define como uno menos la probabilidad de cometer un error tipo II (es decir, si la probabilidad de cometer un error tipo II es β , potencia = $1 - \beta$, o potencia es la probabilidad de que se evite un error tipo II). Los estudios con altos niveles de potencia estadística rara vez no podrán detectar los efectos de los tratamientos. Se asume que la mayoría de los tratamientos tienen al menos algún efecto, entonces la potencia estadística de un estudio se traduce en la probabilidad de que el estudio llegue a la conclusión correcta (Murphy y col., 2014)

El tamaño de un error de tipo I se define como la probabilidad de cometer ese error y, de manera similar, el tamaño de un error de tipo II es la probabilidad de que se cometa

este error. Si la distribución de la que se obtuvo la muestra está parametrizada por θ , donde $\theta \in \Theta$, entonces asociada con cualquier prueba, es una función de potencia. Sea Υ una prueba de la hipótesis nula \mathcal{H}_0 .

Definición. La función de potencia de la prueba Υ , denotada por $\pi_{\Upsilon}(\theta)$, se define como la probabilidad de que \mathcal{H}_0 sea rechazada, cuando la distribución de la que se obtuvo la muestra fue parametrizada por θ (Mood y col., 1974).

Note que, $\pi_{\Upsilon}(\theta) = \mathcal{P}_{\theta}[\text{rechazar } \mathcal{H}_0]$, donde θ es el valor real del parámetro. Si Υ es una prueba no aleatoria, entonces $\pi_{\Upsilon}(\theta) = \mathcal{P}_{\theta}[(X_1, \dots, X_n) \in \mathcal{C}_{\Upsilon}]$, donde \mathcal{C}_{Υ} es la región crítica asociada con la prueba Υ . Si Υ es una prueba aleatoria con función crítica $\Psi_{\Upsilon}(\cdot, \dots, \cdot)$, entonces

$$\begin{aligned} \pi_{\Upsilon}(\theta) &= \mathcal{P}_{\theta}[\text{rechazar } \mathcal{H}_0] \\ &= \int \cdots \int \mathcal{P}_{\theta}[\text{rechazar } \mathcal{H}_0 \mid x_1, \dots, x_n] f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) \prod_{i=1}^n dx_i \\ &= \int \cdots \int \Psi_{\Upsilon}(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) \prod_{i=1}^n dx_i \\ &= \varepsilon_{\theta}[\Psi_{\Upsilon}(X_1, \dots, X_n)]. \end{aligned}$$

Por lo general, será nuestro estándar al evaluar la bondad de una prueba o al comparar dos pruebas en juego. Una función de potencia ideal, es una función que es cero para los θ correspondientes a la hipótesis nula y es la unidad para los θ correspondientes a la hipótesis alternativa. La idea es que, no se desea rechazar \mathcal{H}_0 si \mathcal{H}_0 es verdadera y se desea rechazar \mathcal{H}_0 cuando \mathcal{H}_0 es falsa.

Comúnmente en agronomía se prueba la hipótesis de que no existe diferencia entre tratamientos, que no existen efectos, o que no existe correlación entre variables, por mencionar algunas, lo que se hace es probar hipótesis nulas, de aquí el poder llegar a cometer un error tipo I. De una manera más reducida, se puede ver investigadores que no puedan rechazar \mathcal{H}_0 , cuando en realidad si existe un efecto, es decir, llegar a cometer un error tipo II.

Respecto al análisis estadístico de la potencia (Cohen, 1988; Thiemann y Kraemer, 1987; Lipsey, 1990; Murphy y col., 2014) dicen que, los estudios con muy poco poder estadístico con frecuencia pueden llevar a conclusiones erróneas. En particular, con frecuencia conducen

a la conclusión incorrecta de que los hallazgos informados en un estudio en particular no son ciertos en la población en general.

Un objetivo de la teoría estadística es poder estimar la cantidad de error de muestreo, para poder saber la cantidad de riesgo involucrado en el uso de un procedimiento particular. Las pruebas de significancia nos ayudan como investigadores a obtener conclusiones, si realmente los resultados del estudio de investigación realizado representan a la población, o se deben a un error de muestreo.

Dentro de los elementos más importantes que determinan la potencia de un experimento, se encuentran la varianza, el tamaño del efecto, nivel de significancia, número de tratamientos y número de repeticiones. No tomar estos factores en consideración, y el número apropiado de éstos, al momento de realizar una investigación, pudiera resultar en conclusiones sin un sustento aceptable.

Al buscar una buena prueba, es común restringir la consideración a las pruebas que controlan la probabilidad de error de tipo I en un nivel específico. Dentro de esta clase de pruebas, buscamos pruebas que tengan una probabilidad de error de tipo II que sea lo más pequeño posible (Casella y Berger, 2001).

2.2.3. Tamaño de la prueba

Sea Υ una prueba de hipótesis $\mathcal{H}_0 : \theta \in \Theta_0$, donde $\Theta_0 \subset \Theta$; esto es, Θ_0 es un subconjunto del espacio de parámetros Θ . El tamaño α de la prueba Υ de \mathcal{H}_0 esta definido para ser

$$\sup_{\theta \in \Theta_0} [\pi_{\Upsilon}(\theta)] = \alpha$$

2.2.4. Nivel de significancia

Sea Υ una prueba de hipótesis $\mathcal{H}_0 : \theta \in \Theta_0$, donde $\Theta_0 \subset \Theta$; esto es, Θ_0 es un subconjunto del espacio de parámetros Θ . El nivel α de significancia de la prueba Υ de \mathcal{H}_0 esta definido para ser

$$\sup_{\theta \in \Theta_0} [\pi_{\Upsilon}(\theta)] \leq \alpha$$

Note que, la diferencia de los dos conceptos es que el tamaño de la prueba es el valor del

error tipo I que se comete y el nivel de significancia es el nivel máximo de error tipo I que se esta dispuesto a cometer. De acuerdo con las definiciones, el conjunto de pruebas de nivel α contiene el conjunto de pruebas de tamaño α .

2.2.5. Tamaño del efecto

El tamaño del efecto es un concepto clave en el análisis estadístico de potencia (Cohen, 1988; Rosenthal, 1993; Tatsuoka, 1993). Un índice estandarizado del impacto real de los tratamientos en la variable dependiente, es lo que podemos obtener de las medidas del tamaño del efecto en el nivel más simple, Murphy y col. (2014) nos presenta la diferencia de medias estandarizada, d , definida como $d = \frac{(M_t - M_c)}{SD}$, donde M_t y M_c son las medias del grupo de tratamiento y control, respectivamente, y SD es la desviación estándar agrupada. Al expresar la diferencia en las medias grupales en unidades de desviación estándar, el estadístico d proporciona una métrica simple que permite la comparación de los efectos del tratamiento de diferentes estudios, áreas o investigaciones, etc., sin tener que realizar un seguimiento de las unidades de medida utilizadas en diferentes estudios o áreas de investigación.

La tabla 2.2 presentada por Murphy y col. (2014) que permite realizar análisis en términos de estimaciones del tamaño del efecto en lugar de en términos de sus equivalentes F es

Tabla 2.2. Algunas convenciones para definir al tamaño del efecto.

	PV	r	d	f^2	Probabilidad de un valor mayor en el grupo de tratamientos
Efectos pequeños	.01	.10	.20	.02	.56
Efectos medianos	.10	.30	.50	.15	.64
Efectos grandes	.25	.50	.80	.35	.71

Nota: Cohen's $f^2 = R^2/(1 - R^2) = \eta^2/(1 - \eta^2) = PV/(1 - PV)$, donde $\eta^2 = SC_{Tratamientos}/SC_{Total}$, datos de (Cohen, 1988; Grissom, 1994; Murphy y col., 2014).

donde PV es la estimación de la proporción de varianza total en la variable dependiente que se explica por los tratamientos, r el coeficiente de correlación, d la diferencia de medias estandarizada y f^2 es el cuadrado del estimador del tamaño del efecto f presentado por (Cohen, 1988). El porcentaje de varianza (PV) asociado con cada efecto en un modelo lineal proporciona una medida muy general de si los efectos del tratamiento son grandes o pequeños (es decir, si explican una gran parte de la varianza en la variable dependiente o solo una

pequeña cantidad), y el valor de PV está estrechamente relacionado con F . Se prefiere el término más general PV , porque se refiere al índice general de los efectos de los tratamientos.

Cuanto mayor sea el efecto, mayor será el parámetro de no centralidad (δ) y mayor el valor esperado de F . Cuanto mayor sea la F , más probable es que \mathcal{H}_0 sea rechazada. Por lo tanto, en igualdad de condiciones, cuanto mayor sea la no centralidad (es decir, cuanto mayor sea el efecto o mayor la N), mayor será la potencia.

Note que la nomenclatura utilizada de efecto “pequeño”, “mediano” o “grande” no tiene relación con su importancia, podemos tener ejemplos en donde los efectos de las diferencias son fácilmente detectables y esto no resta importancia.

Por lo regular, para probar hipótesis estadísticas de los trabajos de investigación en agronomía, se fijan los niveles de α en 0.05 o 0.01, lo cuál nos ayuda a minimizar el error tipo I pero tiene un impacto en la potencia estadística, por lo que, siempre habrá una compensación entre los dos tipos de errores, endurecer la prueba, es decir, hacer que sea muy difícil rechazar \mathcal{H}_0 y minimiza el error tipo I, pero aumentará el error tipo II.

Numerosos autores han señalado que los procedimientos para controlar o minimizar los errores tipo I pueden reducir sustancialmente a la potencia estadística y pueden causar más problemas (es decir, errores tipo II) de los que resuelven ([Sedlmeier y Gigerenzer, 1992](#); [Cohen, 1994](#); [Murphy y col., 2014](#)).

La importancia del tamaño del efecto radica en el valor de la potencia, ya que esta estrechamente ligada a ésta, se puede ver en el caso de no tener suficiente potencia de prueba en alguna investigación, esto ocasiona que, al tener tamaños de efecto pequeños es más probable que no se detecten diferencias en las medias de tratamiento, a diferencia de cuando tenemos una prueba con una potencia aceptable, ésta podrá detectar las diferencias en las medias, sin importar el tamaño del efecto.

2.3. Modelo lineal general (LM)

El enfoque más ampliamente utilizado para experimentos, es el uso de modelos lineales. Un modelo lineal se puede expresar en forma matricial como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.6}$$

donde

\mathbf{y} : es un vector de dimensiones $n \times 1$, donde n es el número de observaciones en el conjunto de datos.

\mathbf{X} : es una matriz de $n \times p$, donde p es el número de parámetros a estimar.

$\boldsymbol{\beta}$: es un vector de coeficientes de regresión de dimensión $p \times 1$.

$\boldsymbol{\epsilon}$: es un vector de dimensiones $n \times 1$.

Donde el vector de errores tiene los supuestos

$$\boldsymbol{\epsilon} \rightarrow N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

para un término de error $\boldsymbol{\epsilon}$ que tiene

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \quad \text{y} \quad \mathbf{V} = \text{var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$$

Un conjunto de datos puede estar representado entonces como un vector $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ que se explica por medio de las columnas de lo que se conoce como la matriz modelo

$$\mathbf{X} = x_{ij} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

donde x_{ij} es el valor de la variable explicativa j para la observación i , cuando

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad \text{con} \quad \mathbf{V} = \sigma^2 \mathbf{I}.$$

2.3.1. Supuestos

Los supuestos de un LM son:

1. Independencia

Dentro de los supuestos que debemos cumplir esta la independencia, la cual aumenta la probabilidad de que se cumpla al momento de generar un buen diseño, es decir, antes de llevar a cabo el experimento, una parte fundamental para que se tenga éxito

en la investigación, es hacer uso del azar, esto es, no influir en el experimento ya sea de manera directa o indirecta. Un requisito necesario para la precisión es asignar de manera aleatoria los tratamientos a las UE; esto es importante cuando el cálculo y la interpretación de las estadísticas F (y t) son una parte importante del análisis, lo cual, debe de ser para cualquier experimento, un procedimiento razonable.

Para asegurar la validez de la estimación del error experimental nos basamos en la aleatorización, la cual, conduce a una estimación no sesgada de la varianza, así como a una estimación no sesgada de las diferencias de tratamiento, es decir, estimaciones que están libres de diferencias sistemáticas debidas a una variación no controlada de otro modo.

2. Homogeneidad de varianzas

El segundo supuesto es que, aunque los tratamientos pueden diferir en los rendimientos promedio, la variación del rendimiento entre las UE debería ser similar para todos los tratamientos en el ensayo, lo que conocemos como homogeneidad de varianzas.

3. Normalidad

Para hacer uso de los LM se supone que las variables aleatorias tienen una distribución Normal dentro de cada tratamiento.

Algunas suposiciones son más importantes que otras porque algunas violaciones son más serias debido a que pueden llevar a conclusiones engañosas ([Faraway, 2015](#))

1. La forma sistemática del modelo. Si se equivoca gravemente, las predicciones serán inexactas y cualquier explicación de la relación entre las variables puede estar sesgada de manera engañosa.
2. Dependencia de errores. La presencia de una fuerte dependencia significa que hay menos información en los datos de lo que sugiere el tamaño de la muestra. Además, existe el riesgo de que se introduzca erróneamente componentes sistemáticos en el modelo en un intento de lidiar con una dependencia insospechada en los errores.
3. Varianza no constante. No abordar esta violación de los supuestos del modelo lineal puede resultar en inferencias inexactas. En particular, es posible que la incertidumbre de la predicción no se cuantifique correctamente.

4. Normalidad. Para grandes conjuntos de datos, la inferencia será bastante robusta a una falta de normalidad, ya que el teorema del límite central significará que las aproximaciones tenderán a ser adecuadas. A menos que el tamaño de la muestra sea bastante pequeño o los errores muy anormales, esta suposición no es crucial para el éxito.

2.3.2. Modelo y Análisis de Varianza (ANOVA)

El modelo estadístico lineal se usa cotidianamente en el proceso de generación de conocimiento, es una de las metodologías estadísticas más utilizadas en la modelización y el análisis de datos de todo tipo. A los investigadores les interesa saber cómo el comportamiento de un conjunto de variables impacta sobre otra, no solo desde el punto de la influencia o grado de asociación, sino describir la posible relación funcional entre las mismas.

El objetivo de cualquier pregunta científica es el poder describir o descubrir relaciones entre variables bajo condiciones controladas o reales para predecir eventos futuros, controlar las respuestas, probar algunas hipótesis, por mencionar algunas.

El diseño experimental es un proceso de prueba o investigación de ensayo(s) que busca(n) la verdad, utilizando un diseño o modelo matemático, que nos permita comprobar una hipótesis (Condo y Pazmiño, 2015). Un investigador tiene la necesidad de un análisis estadístico para fundamentar una base objetiva de evaluación, el diseño experimental involucra determinar la forma en la que los niveles de los factores o tratamientos son asignados a las unidades experimentales (UE), la elección del tamaño muestral y la disposición. La idea es aplicar tratamientos a las UE, una o más variables de interés que son manejadas de manera predeterminada, los experimentos realizados en el campo suponen variables no controladas por el experimentador.

Un objetivo principal del análisis de datos en diseños de experimentos es cuantificar y evaluar la importancia de las posibles fuentes de variación. Montgomery (2004) y Lawson (2014) nos dicen que para probar la hipótesis nula de que los efectos de dos o más tratamientos son iguales, se establece el ANOVA, en donde uno de sus modelos estadísticos más simples es el Diseño Experimental Completamente al Azar (DECA), el modelo matemático para las observaciones del experimento se puede representar como

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{2.7}$$

donde:

y_{ij} : es la variable aleatoria correspondiente al tratamiento i en su repetición j .

μ : es la media general.

τ_i : es el efecto del i -ésimo tratamiento $i = 1, \dots, t$.

ϵ_{ij} : es el error aleatorio correspondiente a todas las demás fuentes de variabilidad del experimento, donde $\epsilon_{ij} \sim NIID(0, \sigma^2)$.

Esto implica que $Y_{ij} \sim NI(\mu + \tau_i, \sigma^2)$. El modelo 2.7 nos dice que en la variable respuesta, existe un efecto de la media general y las fuentes de variabilidad son, los tratamientos y el error que emana de la variabilidad del experimento.

Suponga que $j = 2$, y que

$$X_{ij} = \begin{cases} 1 & \text{si es del tratamiento } i \\ 0 & \text{de otra forma} \end{cases}$$

de forma matricial, el modelo se puede ver como

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{t1} \\ y_{t2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_t \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix}.$$

Note que, la columna uno de la matriz \mathbf{X} es una combinación lineal de las demás columnas, por lo que se dice que el modelo esta sobreparametrizado. Habiendo formado una matriz \mathbf{X} del modelo y una \mathbf{y} observada, para obtener estimaciones $\hat{\boldsymbol{\beta}}$ de los parámetros y valores ajustados $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, el método más utilizado para LM es la estimación de mínimos cuadrados, el cuál se obtiene como

$$\hat{\boldsymbol{\beta}} = \text{mín} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right)^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right) \quad \text{sujeto a} \quad \sum \tau_i = 0.$$

Éstas estimaciones tienen la propiedad de que son insesgadas, la varianza dentro de los

tratamientos se estima a partir del cuadrado medio del error (CM_{Error}) en el ANOVA, donde la diferencia entre las medias de tratamientos se evalúa a partir del cuadrado medio de los tratamientos ($CM_{Tratamientos}$) en relación con el CM_{Error} ; y el CM_{Error} es un estimador insesgado de la varianza aleatoria de los errores de las UE, donde

$$SC_{Tratamientos} = \sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \quad , \quad SC_{Error} = \sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

ademas

$$GL_{Tratamientos} = t - 1 \quad , \quad GL_{Error} = n - t,$$

y

$$CM_{Tratamientos} = \frac{SC_{Tratamientos}}{GL_{Tratamientos}} \quad , \quad CM_{Error} = \frac{SC_{Error}}{GL_{Error}},$$

donde n_i es el tamaño de la muestra o las repeticiones observadas en el tratamiento i .

Definición. Sea U una variable aleatoria chi-cuadrada con m grados de libertad; sea V una variable aleatoria chi-cuadrada con n grados de libertad, y sea U y V independientes. Entonces la variable aleatoria

$$X = \frac{U/m}{V/n}, \quad (2.8)$$

se distribuye como F con m y n grados de libertad.

Tomando en cuenta la definición de la distribución F de 2.8, y puesto que los errores ε_{ij} siguen una distribución Normal e independiente con media cero y varianza σ^2 , las observaciones y_{ij} tienen una distribución Normal e independiente con media $\mu + \tau_i$ y varianza σ^2 , por lo tanto, la SC_{Error}/σ^2 es una variable chi-cuadrada con $n - t$ grados de libertad y $SC_{Tratamientos}/\sigma^2$ es una variable aleatoria chi-cuadrada con $t - 1$ grados de libertad, si la hipótesis nula $\tau_i = 0$ para todo $i = 1, 2, \dots, t$ es cierta, cuando se quiere llevar a cabo una prueba formal de hipótesis de que no haya efecto de tratamientos.

Por lo tanto, el cociente

$$F_0 = \frac{SC_{Tratamientos}/(t - 1)}{SC_{Error}/(n - t)} = \frac{CM_{Tratamientos}}{CM_{Error}} \quad (2.9)$$

se distribuye F con $t - 1$ y $n - t$ grados de libertad y se puede utilizar F_0 para probar la

igualdad de las t medias de los tratamientos

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_t \quad \text{vs} \quad \mathcal{H}_a : \exists \text{ un } i \text{ diferente de } j \text{ tal que } \mu_i \neq \mu_j,$$

tomando en cuenta que $\mu_i = \mu + \tau_i$, de manera equivalente

$$\mathcal{H}_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0 \quad \text{vs} \quad \mathcal{H}_a : \text{al menos un } \tau_i \neq 0.$$

La regla de decisión para esta prueba de hipótesis, con un nivel de significancia α es

$$\text{Rechazar } \mathcal{H}_0 \text{ si } F_0 > F_{\alpha, t-1, n-t}.$$

Las estadísticas más utilizadas en las ciencias sociales y del comportamiento se interpretan en términos de, o se traducen fácilmente, en la estadística F . El modelo de análisis de potencia utiliza la distribución F no central para estimar la potencia de una amplia gama de estadísticas ([Patnaik, 1949](#)).

Ésta F no central representa la distribución de resultados que se esperaría encontrar en cualquier estudio en particular (dado un tamaño de efecto); el grado de no centralidad (δ) es una función directa del tamaño del efecto de interés. La potencia estadística del estudio es simplemente la proporción de ésta distribución F no central que se encuentra por encima de cualquier criterio utilizado para definir la significancia estadística.

Retomando el concepto de la probabilidad del error tipo II, tenemos que

$$\begin{aligned} \beta &= 1 - P \{ \text{Rechazar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ es falsa} \} \\ &= 1 - P \{ F_0 > F_{\alpha, t-1, n-t} \mid \mathcal{H}_0 \text{ es falsa} \} \end{aligned} \tag{2.10}$$

ahora, para evaluar 2.10, debemos de considerar la distribución del estadístico de prueba F_0 , que cuando \mathcal{H}_0 es falsa, se distribuye como una variable aleatoria F no central con $t - 1$ y $n - t$ grados de libertad y parámetro de no centralidad δ . Note que, si $\delta = 0$, la distribución F no central se convierte en la distribución F central que se conoce.

El parámetro de no centralidad [Torres y Seguí \(2001\)](#) lo describen como

$$\delta = \frac{SC_{\text{Tratamientos}}}{CM_{\text{Error}}}, \tag{2.11}$$

asociando δ con el concepto de PV , en el que solo se está probando un efecto, y se supone que la varianza se debe a los efectos de los tratamientos o al error, se puede hacer la estimación mediante la siguiente ecuación

$$\delta = GL_{Error} \left[\frac{PV}{1 - PV} \right],$$

además, [Montgomery \(2004\)](#) nos dice que (Φ) es un parámetro que refleja la medida en que la hipótesis nula es falsa, para poder comparar con la potencia de la prueba, donde

$$\Phi^2 = \frac{n \sum_{i=1}^t \tau_i^2}{t\sigma^2}. \quad (2.12)$$

Como se puede ver, la potencia se relaciona directamente con la magnitud del efecto, el número de tratamientos, cantidad de repeticiones en el estudio, media y varianza, los cuales fueron los factores utilizados en este estudio.

2.4. Alternativas de análisis

Al momento de estar trabajando con variables respuesta de conteos, y se parte del hecho de que ésta tiene distribución Poisson, y se analizan suponiendo que se distribuye normal, los supuestos fallan, esto nos lleva a que, la capacidad de detectar efectos reales se altera y la interpretación de los datos puede ser errónea, es por esto que se deben de tomar en cuenta otras alternativas de análisis.

2.4.1. Transformaciones

Las transformaciones de la respuesta y/o los predictores pueden mejorar el ajuste y corregir las violaciones de los supuestos del modelo, como la no constante varianza del error. Cuando se observan problemas en los gráficos de diagnóstico, se sugiere alguna modificación del modelo. Si se observa alguna no linealidad, quizás junto con una varianza no constante, se debe considerar una transformación de las variables.

Un aspecto importante del análisis de datos experimentales es el de la escala de medición, debido a que pueden surgir problemas debido a la falta de aditividad de los efectos unitarios y del tratamiento y/o la falta de constancia de las variaciones. Ambos fenómenos están posiblemente relacionados y pueden resolverse mediante la transformación de los datos a una

escala más apropiada.

Es intrínseco en los análisis de datos experimentales que la aditividad se mantiene; de lo contrario, diferentes experimentos conducirán a la existencia de interacción entre experimentos y tratamientos. Por supuesto, no es posible establecer para qué escala de medición se cumple la aditividad, pero es plausible asociar la no adición con la no constancia de las varianzas, y eso a menudo puede eliminarse mediante una transformación adecuada de las observaciones.

Se demostró que muchas transformaciones, para el supuesto de homogeneidad de varianzas, producen distribuciones simétricas simultáneamente que se aproximan mejor a una distribución Normal. Estas transformaciones tienen una larga historia en el análisis de la varianza (Curtiss, 1943; Fisher, 1954). Una de las ventajas de las transformaciones reside en el uso de modelos lineales de teoría Normal para los cuales las estadísticas inferenciales tienen distribuciones conocidas (Gbur y col., 2012).

Hinkelmann y Kempthorne (1994) comentan que existen diseños experimentales donde la homogeneidad de varianzas no es posible, como al momento de comparar diferentes nutrientes con respecto a la tasa de germinación de ciertas semillas, tenemos n semillas en cada maceta (UE) y para cada semilla la observación es $y = 1$ si la semilla germina e $y = 0$ si no germina. Si para el i -ésimo tratamiento la probabilidad (tasa) de germinación es P , entonces

$$\begin{aligned} E(y_{ij}) &= p_i \\ \text{var}(y_{ij}) &= p_i(1 - p_i) \end{aligned}$$

y

$$\begin{aligned} E(\bar{y}_{ij}) &= p_i \\ \text{var}(\bar{y}_{ij}) &= \frac{p_i(1-p_i)}{n} \end{aligned}$$

Por tanto, si hay diferencias entre los tratamientos, entonces no tenemos constancia de varianzas. Este es un ejemplo donde la varianza es una función de la media de las observaciones. De manera más general, se puede expresar:

Si y es la observación con $E(y) = \mu^*$, entonces

$$\text{var}(y) = g^2(\mu^*) \tag{2.13}$$

donde $g^2(\mu^*)$ es alguna función de μ^* , note que en el ejemplo propuesto $\mu^* = \mu + \tau_i$ si la observación y se obtiene para el i -ésimo tratamiento y que por esta razón los $g^2(\mu^*)$ en 2.13 son posiblemente diferentes. Para igualar o estabilizar la varianza entre tratamientos usamos una transformación de las observaciones y , digamos $f(y)$, tal que

$$\text{var}[f(y)] = \text{constante} = c^2, \quad \text{por decir.} \quad (2.14)$$

Para determinar una transformación adecuada usamos la expansión de la serie de Taylor de $f(y)$ alrededor de μ^* , esto también se conoce como el método de diferenciales estadísticos o el método delta, y se escribe

$$f(y) \cong f(\mu^*) + f'(\mu^*)(y - \mu^*) + \text{recordatorio} \quad (2.15)$$

Tomando el valor esperado de los dos lados de 2.15, tenemos que

$$E[f(y)] \cong f(\mu^*)$$

y por lo tanto, usando 2.13 y 2.15,

$$\text{var}[f(y)] \cong [f'(\mu^*)]^2 g^2(\mu^*) \quad (2.16)$$

Se deduce entonces de 2.14 y 2.16 que

$$f'(\mu^*) = \frac{c}{g(\mu^*)} \quad \text{o} \quad f(\mu^*) = \int \frac{c}{g(\mu^*)} d\mu^* \quad (2.17)$$

Para las observaciones donde 2.13 se cumple (al menos aproximadamente) debido a consideraciones teóricas o debido a evidencia empírica (trazando los residuos, por ejemplo), podemos entonces determinar una transformación apropiada $f(y)$ a partir de 2.17. En la tabla 2.3 se dan ejemplos de algunas transformaciones conocidas presentadas por [Hinkelmann y Kempthorne \(1994\)](#).

[Box y Cox \(1964\)](#) propusieron una clase general de transformaciones que incluyen como casos especiales a ([Snedecor y Cochran, 1989](#)) y ([Steel y col., 1997](#)), los cuales abordaron el problema general de la no normalidad mediante el uso de transformaciones donde el objetivo final de éstas era permitir al investigador obtener análisis aproximados utilizando los métodos estándar de la teoría normal.

Tabla 2.3. Algunas transformaciones conocidas presentadas por [Hinkelmann y Kempthorne \(1994\)](#).

Número	$g^2(\mu^*)$	y	$f(y)$	$\text{var}[f(y)]$	Distribución
1	$\left. \begin{matrix} \mu^* \\ \lambda^2 \mu^{*2} \end{matrix} \right\}$	Conteos	\sqrt{y}	$\left\{ \begin{matrix} .25 \\ .25\lambda^2 \end{matrix} \right.$	Poisson Empírica
2	$\lambda^2 \mu^{*2}$	y_{ij}	$\log y$	λ^2	Empírica
3	$\frac{2\mu^{*2}}{n-1}$	s_{ij}^2	$\log y$	$\frac{2}{n-1}$	Varianza muestral
4	$\frac{\mu^*(1-\mu^*)}{n}$	Número de éxitos	$\arcsen \sqrt{\frac{y}{n}}$	$\left\{ \begin{matrix} 821/n(\text{grados}) \\ .25/n(\text{radianes}) \end{matrix} \right.$	Binomial
5	$\lambda\mu^*(1-\mu^*)$	y_{ij}	$\arcsen \sqrt{y}$	0.25λ (radianes)	Empírica
6	$\lambda^2 \mu^{*2} (1-\mu^*)^2$	y_{ij}	$\log \frac{y}{1-y}$	λ^2	Empírica
7	$\frac{(1-\mu^*)^2}{n-1}$	r_{ij}	$\frac{1}{2} \log \frac{1+y}{1-y}$	$\frac{1}{n-3}$	Correlación muestral
8	$\mu^* + \lambda^2 \mu^{*2}$	Conteos	$\frac{1}{\lambda} \arcsen(\lambda\sqrt{y})$.25	Empírica

Transformaciones de potencia

Para situaciones donde no existe necesariamente una relación entre la media y la varianza, [Box y Cox \(1964\)](#) han propuesto una familia paramétrica de transformaciones:

$$f(y) \equiv \begin{cases} y(\lambda) = \frac{(y^\lambda - 1)}{\lambda} & (\lambda \neq 0) \\ y(0) = \log y \end{cases} \quad (2.18)$$

Debido a la forma de las transformaciones (2.18), también se les conoce como transformaciones de potencia. La idea general es transformar λ de la información y después usar $y(\hat{\lambda})$ como la transformación actual donde $\hat{\lambda}$ es la estimación de λ . Dado que la escala de la observación transformada depende de λ , esto es, $\hat{\lambda}$, [Box y Cox \(1964\)](#) sugirieron utilizar una transformación normalizada.

$$z(\lambda) = \frac{(y^\lambda - 1)}{\lambda y^{\lambda-1}} (\lambda \neq 0)$$

$$z(0) = \dot{y} \log \dot{y}$$

en cambio, donde \dot{y} es la media geométrica de las observaciones y .

Dado que el objetivo principal de las transformaciones (2.18) es lograr la normalidad, el estimador para λ se obtiene asumiendo una distribución normal multivariada para $y(\lambda)$ con una varianza constante, el segundo objetivo, para un LM, el tercer objetivo. Con estos supuestos, es decir, objetivos, $\hat{\lambda}$ puede ser obtenida usando la teoría de MV. También, los niveles de confianza aproximados para λ pueden ser obtenidos dando al usuario una opción más amplia de λ que puede ayudar a interpretar la transformación, es decir, en lugar de usar, por ejemplo $\hat{\lambda} = -.9$ más adecuado, y quizás plausible, la opción puede ser usar $\hat{\lambda} = -1$ ([Hinkelmann y Kempthorne, 1994](#)).

2.4.2. Modelos lineales generalizados (GLM)

Una pregunta que se hacía con frecuencia es ¿por qué se necesita un modelo lineal generalizado?, si en cambio, se pudiera hacer una transformación de la variable respuesta para que los datos transformados se distribuyan aproximadamente de manera normal.

A veces no se reconoce que el uso de transformaciones cambie el modelo en consideración. [McArdle y Anderson \(2004\)](#) discutieron el ejemplo de la transformación de log, que a menudo se utiliza para distribuciones sesgadas para obtener simetría. En una distribución simétrica la media y la mediana son las mismas, de modo que un intervalo de confianza para la media de los datos transformados de log, también es un intervalo de confianza para la mediana. La transformación inversa de los puntos finales del intervalo de confianza produce un intervalo de confianza para la mediana en la escala original porque la monotonidad del algoritmo y las funciones exponenciales preservan el orden de los datos. Sin embargo, la media de los datos en la escala de log no es igual a el log de la media de los datos originales. Por lo tanto, el intervalo de confianza transformado no proporciona información sobre la media en la escala original. Los GLM evitan estos problemas, porque los datos no se transforman; en cambio, una función de las medias se modela como una combinación lineal de las variables explicativas ([Gbur y col., 2012](#)).

Una desventaja de usar transformaciones es que la elección de transformación es subjetiva. [Mahmud y col. \(2006\)](#) demostraron que probar varias transformaciones en los datos y luego elegir una a posteriori infla la tasa de error tipo I, lo que hace que sea más fácilmente rechazar falsamente la hipótesis nula. Por lo tanto, las transformaciones pueden ser problemáticas cuando una elección particular no está predeterminada por otras consideraciones.

Los GLM permiten que la media de una población dependa de un predictor lineal a través de una función de enlace (link function) de tipo no lineal y que la distribución probabilística de la variable respuesta sea cualquiera de la familia exponencial. De este modo, las variables explicativas pueden ser cualquier combinación de variables continuas, variables de clasificación y sus interacciones. La estimación de parámetros en estos modelos se realiza mediante procedimientos de MV, Manual de usuario de SAS/IML 9.3 ([Instituto SAS, 2011](#)).

Cuando se está haciendo investigación en las ciencias agrícolas por lo regular se manejan como tratamientos y error a la suma de “componentes” que explican a la variable respuesta y que se generan de manera aleatoria, los cuales nos ayudan a tratar de hacer inferencia

sobre alguna población en particular; para diseños más elaborados se pueden incluir bloques y covariables.

Utilizando el concepto de cuasi-verosimilitud, (Nelder y Wedderburn, 1972) extendieron la aplicabilidad de los GLM utilizando cuasi-verosimilitud a situaciones donde no se puede especificar la distribución exactamente, se puede especificar la forma de la relación media / varianza, y si las observaciones son independientes o no están correlacionadas, se puede ajustar el modelo y obtener resultados no iguales, pero similares a los que se habrían obtenido si la distribución hubiera sido conocida.

2.4.2.1. Componentes

Los GLM extienden los LM de la teoría normal a variables de respuesta cuyas distribuciones pertenecen a la familia exponencial o pueden caracterizarse por una cuasi-verosimilitud. Esta clase de modelos incluye análisis de efectos fijos de varianza, así como regresión y análisis de modelos de covarianza que no contienen efectos aleatorios. Gbur y col. (2012) nos dicen que un GLM consta de tres componentes:

1. Un componente estocástico que define la distribución de probabilidad de la familia exponencial o cuasi-verosimilitud de la variable respuesta.
2. Un predictor lineal que es un componente sistemático que describe el modelo lineal definido por las variables explicativas.
3. Una función de enlace $g(\cdot)$ que relaciona la media de la variable de respuesta con una combinación lineal de las variables explicativas.

Se debe de hacer notar que, un GLM relaciona la media transformada de la variable respuesta con las variables explicativas, pero no implica transformar la variable respuesta en si misma, por lo que los datos permanecen en la escala de medición original, pero la escala del modelo para la media como función lineal esta en una escala diferente (la del enlace). El enfoque es diferente a las transformaciones, donde los datos sufren una transformación de la escala original.

El componente predictor lineal de un GLM crea una relación intrínsecamente lineal entre una función de la media de la respuesta y las variables explicativas. Las funciones de enlace válidas son monótonas y son diferenciables. Si $g(\cdot)$ tiene la misma forma funcional que el parámetro canónico $\theta = \eta(\cdot)$ de la distribución de \mathbf{X} , se denomina enlace canónico (Gbur

y col., 2012).

Mediante los métodos estadísticos existentes, la función liga permite realizar pruebas de hipótesis e inferencias como las pruebas t y las pruebas F en la escala de la liga, que es lineal en las β' s. Dado que la variable de respuesta no se transforma, su distribución supuesta proporciona la base para la estimación e inferencia de MV. Uno de los puntos importantes de la función liga es la monotonicidad, que garantiza que la inversa de la función liga, $g^{-1}(\cdot)$, exista y se puedan transformar los resultados de la escala de la liga a la escala en la que se observó la respuesta \mathbf{X} .

2.4.2.2. Modelos

Familia exponencial de densidades. Una familia de densidades de un parámetro $f(\cdot; \theta)$ (θ es unidimensional) que se pueda escribir de la forma:

$$f(x; \theta) = a(\theta)b(x)\exp[c(\theta)d(x)], \quad (2.19)$$

para $-\infty < x < \infty$ para todo $\theta \in \Omega$, pertenece a la familia exponencial.

La cual se puede generalizar a k parámetros de la siguiente forma.

Familia exponencial de k -parámetros. Una familia de densidades $f(\cdot; \theta_1, \dots, \theta_k)$ que puede ser expresada como

$$f(x; \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k)b(x)\exp\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k)d_j(x) \quad (2.20)$$

está definida que pertenece a la familia exponencial (Mood y col., 1974).

2.4.2.3. Máxima verosimilitud (MV)

Se pueden utilizar diferentes enfoques para la estimación de parámetros que se desconocen de una supuesta distribución de probabilidad, dentro de los cuales, el de mínimos cuadrados es el más utilizado para los LM, los cuales tienen una respuesta que se distribuye de manera normal, pero cuando la respuesta no tiene éste comportamiento, resulta ser más problemático.

Otro de los enfoques que se utiliza se basa en el concepto de verosimilitud, el cuál es usado ampliamente, a este enfoque se le conoce como MV, el cuál, para n variables aleato-

rias X_1, X_2, \dots, X_n la función de MV se define como la densidad conjunta de las n variables aleatorias, digamos $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$, la cuál es considerada ser una función de θ . En particular, si X_1, \dots, X_n es una muestra aleatoria de la densidad $f(x; \theta)$, entonces la función de MV es $f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta)$.

Para hacer notar que estamos hablando de la función de MV como una función de θ , se usará la notación $L(\theta; x_1, \dots, x_n)$. Lo que se quiere obtener es el valor de θ en Θ , denotado por $\hat{\theta}$, el cuál maximiza la función de verosimilitud.

Definición. Si $\hat{\theta}$ (donde $\hat{\theta} = \hat{\vartheta}(x_1, x_2, \dots, x_n)$ es una función de las observaciones x_1, x_2, \dots, x_n) es el valor de θ en Θ que maximiza $L(\theta)$, entonces $\hat{\Theta} = \hat{\vartheta}(X_1, X_2, \dots, X_n)$ es el estimador de MV de θ .

Los casos más importantes que debemos considerar son aquellos en los cuales X_1, X_2, \dots, X_n son una muestra aleatoria de la densidad $f(x; \theta)$, por lo que la función de MV es

$$L(\theta) = f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta). \quad (2.21)$$

Escrito de otra manera, podemos decir que la función es

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta). \quad (2.22)$$

Muchas funciones de MV satisfacen condiciones de regularidad, por lo que el estimador es la solución a la ecuación

$$\frac{dL(\theta)}{d\theta} = 0. \quad (2.23)$$

A veces es más fácil obtener $\hat{\theta}$ si maximizamos $\ell(\theta) = \log L(\theta)$, dado que la función log es una función monótona creciente, y tienen su máximo en el mismo valor de θ (Mood, 1950).

Si la función de MV contiene k parámetros, esto es, si

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k), \quad (2.24)$$

entonces los estimadores de MV de los parámetros $\theta_1, \theta_2, \dots, \theta_k$ son las variables aleatorias $\hat{\Theta}_1 = \hat{\vartheta}_1(X_1, \dots, X_n), \hat{\Theta}_2 = \hat{\vartheta}_2(X_1, \dots, X_n), \dots, \hat{\Theta}_k = \hat{\vartheta}_k(X_1, \dots, X_n)$, donde $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ son los valores en Θ que maximizan $L(\theta_1, \theta_2, \dots, \theta_k)$.

Si ciertas condiciones de regularidad se satisfacen, el punto donde la verosimilitud es un máximo es la solución de las k ecuaciones

$$\begin{aligned}\frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_1} &= 0 \\ \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_k} &= 0.\end{aligned}\tag{2.25}$$

En este caso, también es más fácil trabajar con logaritmos.

Suponiendo que se conoce la forma funcional para la distribución de \mathbf{X} , encontrar estimadores de MV es un problema de optimización. Las técnicas de cálculo diferencial proporcionan un enfoque general de la solución. En algunos casos, es posible una solución analítica; en otros, se deben emplear algoritmos numéricos iterativos. Dado que una función no negativa y su logaritmo natural se maximizan con los mismos valores de la variable independiente, a menudo es más conveniente algebraicamente encontrar el máximo del logaritmo natural de la función de probabilidad.

2.4.2.4. Inferencia

Los parámetros y las pruebas de hipótesis se estiman a través de MV. Las pruebas de hipótesis se establecen utilizando estadísticas de tipo Wald basadas en los estimadores de MV, los cuales son asintóticamente normales. También se pueden usar pruebas de razón de verosimilitud (RV) además de las pruebas de Wald.

La RV es una técnica para llegar a una buena prueba, la cual, para GLM tienden a tener mejores propiedades de error tipo I que las estadísticas de Wald ([Ives, 2015](#)).

Razón de verosimilitud (RV). Sea X_1, \dots, X_n una muestra aleatoria de, ya sea $f_0(\cdot)$ o $f_1(\cdot)$. Una prueba Υ de $\mathcal{H}_0 : X_i \sim f_0(\cdot)$ contra $\mathcal{H}_a : X_i \sim f_1(\cdot)$ se define como una prueba de razón de verosimilitud simple si Υ se define como

$$\begin{aligned}\text{Rechazar } \mathcal{H}_0 &\quad \text{si } \lambda < k, \\ \text{No rechazar } \mathcal{H}_0 &\quad \text{si } \lambda > k, \\ \text{Rechazar o no rechazar } \mathcal{H}_0 &\quad \text{si } \lambda = k,\end{aligned}$$

donde

$$\lambda = \lambda(x_1, \dots, x_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} = \frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} = \frac{L_0}{L_1} \quad (2.26)$$

y k es una constante no negativa. $L_j = L_j(x_1, \dots, x_n)$ es la función de verosimilitud para muestras de la densidad $f_j(\cdot)$.

Ahora, si se comparan dos hipótesis compuestas, donde asumimos que tenemos una muestra aleatoria de $f(x; \theta), \theta \in \Theta$ y queremos probar $\mathcal{H}_0 : \theta \in \Theta_0$ contra $\mathcal{H}_a : \theta \in \Theta_1$, donde $\Theta_0 \subset \Theta, \Theta_1 \subset \Theta$, y Θ_0 y Θ_1 son disjuntos y sea $L(\theta; x_1, \dots, x_n)$ la función de verosimilitud para una muestra X_1, \dots, X_n que tienen una función de densidad conjunta $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta)$ donde $\theta \in \Theta$, la prueba de RV generalizada esta definida por

$$\lambda = \lambda_n = \lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)} \quad (2.27)$$

Como ejemplo, si asumimos que tenemos k muestras aleatorias independientes, una de cada una de las poblaciones normales, esto es, sea X_{j1}, \dots, X_{jn} , una muestra aleatoria de tamaño n_j de la j -ésima población normal, $j = 1, \dots, k$. Asumiendo que la j -ésima población tiene media μ_j y varianza σ^2 . El objetivo es probar la hipótesis nula que todas las medias de las poblaciones son iguales contra la alternativa que no todas las medias son iguales, la función de verosimilitud esta dada por

$$\begin{aligned} L(\mu_1, \dots, \mu_k, \sigma^2; x_{11}, \dots, x_{1n_1}, \dots, x_{k1}, \dots, x_{kn_k}) \\ &= \prod_{j=1}^k \prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[(x_{ji} - \mu_j)/\sigma]^2} \\ &= (2\pi\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \mu_j)^2 \right], \end{aligned}$$

donde $n = \sum_{j=1}^k n_j$.

El espacio del parámetro Θ es $(k+1)$ -dimensional con coordenadas $(\mu_1, \dots, \mu_k, \sigma^2)$, y Θ_0 , la colección de puntos en el espacio del parámetro correspondiente a la hipótesis nula, es bidimensional con coordenadas (μ, σ^2) , donde $\mu = \mu_1 = \dots = \mu_k$. En Θ los estimadores de

MV de $\mu_1, \dots, \mu_k, \sigma^2$ están dados por

$$\hat{\mu}_j = \bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}, \quad j = 1, \dots, k,$$

y

$$\hat{\sigma}_{\Theta}^2 = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2;$$

por lo tanto,

$$\sup_{\Theta} L = \left[\frac{2\pi \sum_j \sum_i (x_{ji} - \bar{x}_j)^2}{n} \right]^{-n/2} e^{-n/2}.$$

En Θ , los estimadores de MV de μ y σ^2 son

$$\hat{\mu} = \bar{\bar{x}} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji})$$

y

$$\hat{\sigma}_{\Theta_0}^2 = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{\bar{x}})^2$$

por lo tanto

$$\sup_{\Theta_0} L = \left[\frac{2\pi \sum_j \sum_i (x_{ji} - \bar{\bar{x}})^2}{n} \right]^{-n/2} e^{-n/2}.$$

Entonces, tenemos que la RV generalizada es

$$\begin{aligned}
\lambda &= \frac{\sup_{\Theta_0} L}{\sup_{\Theta} L} = \left[\frac{\sum_j \sum_i (x_{ji} - \bar{x})^2}{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2} \right]^{-n/2} \\
&= \left[\frac{\sum_j \sum_i (x_{ji} - \bar{x}_j + \bar{x}_j - \bar{x})^2}{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2} \right]^{-n/2} \\
&= \left[\frac{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2 + \sum_j n_j (\bar{x}_j - \bar{x})^2}{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2} \right]^{-n/2} \\
&= \left[1 + \frac{k-1}{n-k} \frac{\sum_j n_j (\bar{x}_j - \bar{x})^2 / (k-1)}{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2 / (n-k)} \right]^{-n/2}
\end{aligned}$$

Una prueba de RV generalizada viene dada por la siguiente prueba:

$$\text{Rechazar } \mathcal{H}_0 \quad \text{si y solo si } \lambda \leq \lambda_0.$$

Pero $\lambda \leq \lambda_0$ si y solo si

$$r = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 / (k-1)}{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2 / (n-k)} \geq \text{alguna constante, por decir } c. \quad (2.28)$$

La razón r a veces es llamada razón de varianza o razón de F . La constante c es determinada para que la prueba tenga tamaño α ; esto es c se selecciona de manera que

$$P[R \geq c \mid \mathcal{H}_0] = \alpha$$

Note que \bar{X}_j es independiente de $\sum_i (X_{ji} - \bar{X}_j)^2$ y por lo tanto, el numerador de 2.28 es independiente del denominador. También, bajo \mathcal{H}_0 , note que, el numerador dividido entre σ^2 tiene una distribución chi-cuadrada con $k-1$ grados de libertad, y el denominador dividido entre σ^2 tiene también una distribución chi-cuadrada con $n-k$ grados de libertad. Por con-

siguiente, si \mathcal{H}_0 es verdadera, R tiene una distribución F con $k - 1$ y $n - k$ grados de libertad.

El interés sería determinar el tamaño de la prueba para cualquier constante λ_0 o encontrar λ_0 para que la prueba tenga tamaño α , pero la distribución de la RV generalizada no tiene solución. Una prueba aproximada de tamaño α puede obtenerse para n_j grandes.

Sea X_1, \dots, X_n una muestra aleatoria con distribución conjunta $f_{X_1, \dots, X_n}(\cdot, \dots, \cdot; \theta)$, donde $\theta = (\theta_1, \dots, \theta_k)$, que se supone que satisface condiciones de regularidad bastante generales. Suponga que el espacio de parámetros Θ es k -dimensional. Al probar la hipótesis

$$\mathcal{H}_0 : \theta_1 = \theta_1^0, \dots, \theta_r = \theta_r^0, \theta_{r+1}, \dots, \theta_k,$$

donde $\theta_1^0, \dots, \theta_r^0$ son desconocidos y $\theta_{r+1}, \dots, \theta_k$ quedan sin especificar, $-2 \log \Lambda_n$ se distribuye aproximadamente chi-cuadrada con r grados de libertad cuando \mathcal{H}_0 es verdadera y el tamaño de la muestra n es grande (Mood y col., 1974).

Hasta ahora, se ha asumido que $1 \leq r \leq k$ en el teorema anterior. Si $r = k$, entonces todos los parámetros están especificados. El espacio de parámetros Θ es k -dimensional, y dado que \mathcal{H}_0 especifica el valor de r de los componentes de $(\theta_1, \dots, \theta_k)$, la dimensión de Θ_0 es $k - r$. Así, los grados de libertad de la distribución asintótica pueden pensarse de dos formas, primero como el número de parámetros especificados por \mathcal{H}_0 , y segundo, como la diferencia en las dimensiones de Θ y Θ_0 .

Recordando que Λ_n es la variable aleatoria que tiene valores en 2.27, que a su vez es la RV generalizada para una muestra de tamaño n . Θ_0 es el subconjunto de Θ que es especificado por \mathcal{H}_0 . El principio de la RV generalizada dicta que \mathcal{H}_0 se rechaza para λ_n pequeño, pero dado que $-2 \log \lambda_n$ aumenta a medida que λ_n disminuye, una prueba que es equivalente a la prueba de RV generalizada es una que rechaza para $-2 \log \lambda_n$ grande. Ahora, dado que el teorema da una distribución aproximada para los valores $-2 \log \lambda_n$ donde \mathcal{H}_0 es verdadera, una prueba con aproximadamente tamaño α es dada por

$$\text{Rechazar } \mathcal{H}_0 \quad \text{si y solo si} \quad -2 \log \lambda_n > \chi_{1-\alpha}^2(r),$$

donde $\chi_{1-\alpha}^2(r)$ es el $(1 - \alpha)$ cuantil de la distribución chi-cuadrada con r grados de libertad, note que r es el número de componentes del espacio del parámetro que esta especificado por la hipótesis nula.

Al igual que en los LM, como ya hemos comentado, en los GLM también es importante verificar los supuestos del modelo, en donde se pueden utilizar las mismas estadísticas de diagnóstico y verificación de los LM para los GLM. En todos los casos, la verificación del modelo debe abordar la idoneidad del componente estocástico que define la distribución de probabilidad de la variable de respuesta, la función de enlace y la función de varianza. Para hacer una verificación informal de los modelos, se puede realizar el análisis gráfico de los residuales.

Un enfoque para los métodos formales de verificación de modelos incorpora el modelo actual en una clase más amplia de modelos que pueden incluir parámetros adicionales en el predictor lineal, otras distribuciones para la variable de respuesta, diferentes enlaces y/o diferentes funciones de varianza. Se preferiría el modelo actual a otros modelos de la clase si no mejoran apreciablemente el ajuste del modelo en comparación con el modelo actual (McCullagh y Nelder, 1989).

2.4.3. GLM Poisson

La distribución que es más simple para cuando tenemos conteos, colocando su masa en el conjunto de valores enteros no negativos, es la Poisson. Sus probabilidades dependen de un solo parámetro, la media $\mu > 0$ (Agresti, 2015).

En la ecuación (2.3) observamos que la función de masa de probabilidad, $p(y; \mu) = \exp^{-\mu} \mu^y / y!$ para $y = 0, 1, 2, \dots$, esta en la familia de dispersión exponencial con $E(y) = \text{var}(y) = \mu$. La distribución Poisson es unimodal con moda igual a la parte entera de μ . Su falta de simetría es descrita por $E(y - \mu)^3 / \sigma^3 = 1 / \sqrt{\mu}$. Conforme la μ aumenta, la distribución Poisson es menos sesgada y se aproxima a la normalidad, la aproximación siendo bastante buena cuando $\mu > 10$.

La distribución Poisson a menudo se usa para el recuento de eventos que ocurren aleatoriamente en el tiempo o el espacio a una velocidad particular, cuando los resultados en períodos de tiempo o regiones disjuntos son independientes. La Poisson también aplica como una aproximación de la binomial cuando el número de eventos n es grande y π es muy pequeño, con $\mu = n\pi$. Para la binomial, si $n \rightarrow \infty$ y $\pi \rightarrow 0$ tal que $n\pi = \mu$ es fijo, entonces la distribución binomial converge a la Poisson.

2.4.3.1. Estabilización de varianza y mínimos cuadrados con datos de conteo

Sea y_1, \dots, y_n que denota las observaciones independientes de una distribución Poisson, con $\mu_i = E(y_i)$. En el modelado de datos de conteo, podríamos transformar los conteos para que, al menos aproximadamente, la varianza sea constante y los métodos de mínimos cuadrados ordinarios sean válidos. Por medio del método delta, la linearización $g(y) - g(\mu) \approx (y - \mu)g'(\mu)$ implica que la $\text{var}[g(y)] \approx [g'(\mu)]^2 \text{var}(y)$. Si y tiene una distribución Poisson, entonces \sqrt{y} tiene

$$\text{var}(\sqrt{y}) \approx \left(\frac{1}{2\sqrt{\mu}}\right)^2 \mu = \frac{1}{4}$$

La aproximación es mejor para μ grandes, para las cuales \sqrt{y} es más estrechamente lineal en un “vecindario” de μ .

Dado que \sqrt{y} tiene aproximadamente una varianza constante, podemos modelar $\sqrt{y_i}$, $i = 1, \dots, n$, utilizando LM definidos por mínimos cuadrados ordinarios. Sin embargo, el modelo es entonces lineal en $E(\sqrt{y_i})$, no en $E(y_i)$. Además, una relación lineal con el predictor lineal puede ser más pobre para $E(\sqrt{y_i})$ que para $E(y_i)$, $E[\log(y_i)]$, o alguna otra transformación. Es más atractivo utilizar métodos GLM, que aplican una función de enlace a la respuesta media, en lugar de la media a una función de la respuesta.

2.4.3.2. GLM Poisson y Modelos Loglineales

Dado que $\text{var}(y_i) = \mu_i$, las ecuaciones de MV para n observaciones independientes simplifican para una respuesta Poisson con un predictor lineal $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}$ teniendo una función liga g para

$$\sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\mu_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = 0$$

Aunque un GLM puede modelar una media positiva utilizando el enlace de identidad, es más común modelar el log de la media. Al igual que el predictor lineal, la media logarítmica puede tomar cualquier valor real. La media logarítmica es el parámetro natural para la distribución de Poisson, y el enlace log es el enlace canónico para un GLM de Poisson. El modelo loglineal de Poisson es

$$\log \mu_i = \sum_{j=1}^p \beta_j x_{ij}, \quad \text{o} \quad \log \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

En términos de una matriz modelo y los parámetros del modelo. Para $\eta_i = \log \mu_i$, $\partial \mu_i / \partial \eta_i = \mu_i$, por lo que las ecuaciones de MV son

$$\sum_i (y_i - \mu_i) x_{ij} = 0, \quad (2.29)$$

Para un modelo loglineal Poisson, la media satisface la relación exponencial

$$\mu_i = \exp\left(\sum_{j=1}^p \beta_j x_{ij}\right) = (\exp^{\beta_1})^{x_{i1}} \cdots (\exp^{\beta_p})^{x_{ip}}.$$

Un aumento de 1 unidad en x_{ij} tiene un impacto multiplicativo de \exp^{β_j} : la media en $x_{ij} + 1$ equivale a la media en x_{ij} multiplicada por \exp^{β_j} , ajustando por las otras variables explicativas.

2.4.3.3. Ajuste del modelo y bondad de ajuste

A excepción de los modelos simples, como el diseño de una vía o el diseño bidireccional equilibrado, las ecuaciones de probabilidad no tienen una solución de forma cerrada. Sin embargo, la función log-verosimilitud es cóncava, y el método Newton-Raphson (que es equivalente a Fisher para el enlace log canónico) produce valores y estimaciones de los parámetros del modelo correspondientes. La matriz de covarianza asintótica estimada de $\hat{\boldsymbol{\beta}}$ es

$$\text{vâr}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1},$$

donde con el enlace log \mathbf{W} es la matriz diagonal con los elementos $w_i = (\partial \mu_i / \partial \eta_i)^2 / \text{var}(y_i) = \mu_i$.

Tenemos que la devianza de un GLM Poisson es

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right]. \quad (2.30)$$

Cuando un modelo con una función liga log tiene un intercepto, su ecuación de MV implica que $\sum_i \hat{\mu}_i = \sum_i y_i$, y entonces $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_i [y_i \log(y_i / \hat{\mu}_i)]$. Ésto es denotado por G^2 . La estadística correspondiente de Pearson es

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

En algunos casos, podemos usar éstas estadísticas para probar su bondad. Las distribuciones chi-cuadradas asintóticas se producen cuando el número n de observaciones Poisson se fija y sus medias aumentan sin límites. Es más informativo verificar un modelo comparándolo con modelos más complejos (por ejemplo, con términos de interacción) e investigando aspectos particulares de su falta de ajuste. Por ejemplo, podemos verificar que la varianza realmente tenga el mismo orden de magnitud que la media comparando el modelo con un modelo más complejo que no asume esto, como el modelo Binomial negativo.

Agresti (2015) comenta también que podemos buscar observaciones o patrones inusuales en los residuales. Como y_i , éstas tienen distribuciones sesgadas, cada vez menos, a medida que μ_i aumenta. Finalmente, una forma informal de evaluar la suposición de Poisson es comparar la proporción global de la muestra de observaciones (0, 1, 2, ...) con el promedio de las distribuciones de respuesta ajustadas para las n observaciones. A menudo, esto muestra que un modelo de Poisson no permite una variabilidad suficiente, lo que predice cero resultados y resultados relativamente altos.

La devianza (2.30) y la estadística de Pearson para el modelo Poisson para el diseño de una sola vía se simplifican a

$$G^2 = 2 \sum_{i=1}^c \sum_{j=1}^{n_i} y_{ij} \log \left(\frac{y_{ij}}{\bar{y}_i} \right), \quad X^2 = \sum_{i=1}^c \sum_{j=1}^{n_i} \frac{(y_{ij} - \bar{y}_i)^2}{\bar{y}_i}.$$

Para probar bondad de ajuste de este modelo con una relativa grande $\{\bar{y}_i\}$, G^2 y X^2 tienen aproximadamente distribución chi-cuadrada con $df = \sum_i (n_i - 1)$ (Fisher, 1970). Para un grupo simple, Cochran (1954) se refirió a $[\sum_j (y_{1j} - \bar{y}_1)^2] / \bar{y}_1$ como la prueba de la varianza para el ajuste de una distribución Poisson, ya que compara la varianza muestral de los datos con la varianza estimada de Poisson \bar{y}_1 . Ésta teoría asintótica aplica, sin embargo, como $\{\mu_i\}$ va creciendo para $\{n_i\}$ fijos, lo cuál no es realista en la mayoría de las aplicaciones. Para verificar el ajuste, las chi-cuadradas asintóticas generalmente se aplican mejor para comparar el modelo con un modelo más complejo y para comparar el modelo con el modelo nulo como se acaba de describir.

2.4.4. GLM Binomial negativo

Para la distribución de Poisson, la varianza es igual a la media. En la práctica, las observaciones de conteo a menudo exhiben una variabilidad que excede la predicha por el Poisson, a este fenómeno se llama sobredispersión.

2.4.4.1. Sobredispersión para un GLM Poisson

Una razón común para la sobredispersión es la heterogeneidad: a niveles fijos de las variables explicativas, la media varía según los valores de las variables no observadas. Si la varianza es igual a la media cuando se incluyen todas las variables explicativas relevantes, excede la media cuando solo se incluyen algunas. Otra limitación severa de GLM Poisson es que, debido a que la varianza de y debe ser igual a la media, en una media fija, la varianza no puede disminuir a medida que entran en el modelo variables explicativas adicionales.

La sobredispersión no es un problema en los LM que asumen que y se distribuye normalmente, porque esa distribución tiene un parámetro de varianza separado para describir la variabilidad. Para las distribuciones de Poisson y Binomial, sin embargo, la varianza es una función de la media. La sobredispersión es común en el modelado de conteos. Suponga que el modelo para la media tiene la función de enlace correcta y el predictor lineal, pero la distribución de respuesta verdadera tiene más variabilidad que el Poisson. Luego, los estimadores de MV de los parámetros del modelo, suponiendo una respuesta de Poisson, siguen siendo consistentes, convergiendo en probabilidad a los valores de los parámetros, pero los errores estándar son demasiado pequeños. Las extensiones de GLM Poisson que tienen un parámetro adicional representan mejor la sobredispersión.

2.4.4.2. Binomial Negativa como una mezcla Gama de Poisson

Un modelo mixto es una forma flexible de explicar la sobredispersión. En una configuración fija de las variables explicativas realmente observadas, dada la media λ , suponga que la distribución de y es Poisson (λ), pero λ en si varía debido a covariables no medidas. Sea $\mu = E(\lambda)$. Entonces incondicionalmente,

$$E(y) = E[E(y | \lambda)] = E(\lambda) = \mu,$$

$$\text{var}(y) = E[\text{var}(y | \lambda)] + \text{var}[E(y | \lambda)] = E(\lambda) + \text{var}(\lambda) = \mu + \text{var}(\lambda) > \mu.$$

Agresti (2015) nos presenta un ejemplo importante de un modelo mixto para datos de conteo: suponga que (1) dado λ , y tiene una distribución Poisson (λ), y (2) λ tiene una distribución gama, ver ecuación 2.5. Es decir, la función de densidad de probabilidad gama de dos parámetros para y , parametrizada en términos de su media μ y el parámetro de forma $k > 0$, es

$$f(y; k, \mu) = \frac{(k/\mu)^k}{\Gamma(k)} \exp -ky/\mu y^{k-1}, \quad y \geq 0, \quad (2.31)$$

$$\text{para la cuál } E(\lambda) = \mu, \quad \text{var}(\lambda) = \mu^2/k;$$

entonces la desviación estándar es proporcional a la media. Marginalmente, la mezcla gama de las distribuciones de Poisson produce la distribución Binomial negativa para y . Su función de masa de probabilidad es

$$p(y; \mu, k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k}\right)^y \left(\frac{k}{\mu+k}\right)^k \quad y = 0, 1, 2, \dots \quad (2.32)$$

Con k fijo, esta es un miembro de una familia de dispersión exponencial apropiada para variables discretas, con parámetro natural $\log[\mu/(\mu+k)]$.

En la familia Binomial negativa de dos parámetros, sea $\gamma = 1/k$. Entonces y tiene

$$E(y) = \mu, \quad \text{var}(y) = \mu + \gamma\mu^2.$$

El índice $\gamma > 0$ es un tipo de parámetro de dispersión. Cuanto mayor sea el valor de γ , mayor es la sobredispersión en relación con la Poisson. Conforme $\gamma \rightarrow 0$, $\text{var}(y) \rightarrow \mu$, la distribución Binomial negativa converge a la Poisson.

La distribución Binomial negativa tiene un alcance mucho mayor que la Poisson. Por ejemplo, la moda de Poisson es la parte entera de la media y es igual a 0 solo cuando $\mu < 1$. La binomial negativa es también unimodal, pero la moda es 0 cuando $\gamma \geq 1$ y de otro modo, es la parte entera de $\mu(1-\gamma)$. La moda puede ser 0 para cualquier μ .

2.4.4.3. GLMs Binomial negativo

Los GLM Binomiales negativos suelen utilizar el enlace log, como en los modelos loglineales de Poisson, en lugar del enlace canónico. Por simplicidad, dejamos que el parámetro

de dispersión γ sea la misma constante para todas las n observaciones pero tratado como desconocido, como la varianza en los modelos normales. Esto corresponde a un coeficiente de variación constante en la distribución gama mixta, $\sqrt{\text{var}(\lambda)}/E(\lambda) = \sqrt{\gamma}$.

De la ecuación (2.32) expresada en términos de la dispersión del parámetro γ , la función log-verosimilitud para un GLM Binomial negativo con n observaciones independientes es

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) = \sum_{i=1}^n \left[\log \Gamma \left(y_i + \frac{1}{\gamma} \right) - \log \Gamma \left(\frac{1}{\gamma} \right) - \log \Gamma(y_i + 1) \right] \\ + \sum_{i=1}^n \left[y_i \log \left(\frac{\gamma \mu_i}{1 + \gamma \mu_i} \right) - \left(\frac{1}{\gamma} \right) \log(1 + \gamma \mu_i) \right],$$

donde μ_i es una función de $\boldsymbol{\beta}$ a través de $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}$ con la función liga g . La ecuación de verosimilitud obtenida mediante la diferenciación $L(\boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y})$ con respecto a $\boldsymbol{\beta}$ tiene la forma usual para un GLM,

$$\sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\mu_i + \gamma \mu_i^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = 0, \quad j = 1, 2, \dots, p.$$

El log-verosimilitud produce una matriz Hessiana que tiene

$$\frac{\partial^2 L(\boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y})}{\partial \beta_j \partial \gamma} = - \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{(1 + \gamma \mu_i)^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right).$$

Así, $E(\partial^2 L / \partial \beta_j \partial \gamma) = 0$ para cada j , y $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$ son parámetros ortogonales. Entonces $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\gamma}}$ son asintóticamente independientes, y la muestra grande SE para $\hat{\boldsymbol{\beta}}$ es la misma ya sea que γ es conocido o estimado.

El algoritmo de mínimos cuadrados iterativamente ponderado para la puntuación de Fisher se aplica para el ajuste del modelo de MV. La matriz de covarianza estimada de $\hat{\boldsymbol{\beta}}$ es

$$\hat{\text{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1},$$

donde, con la liga log, \mathbf{W} es la matriz diagonal con $w_i = (\partial \mu_i / \partial \eta_i)^2 / \text{var}(y_i) = \mu_i / (1 + \gamma \mu_i)$. La devianza para un GLM Binomial negativo es

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_i \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - \left(y_i + \frac{1}{\hat{\gamma}} \right) \log \left(\frac{1 + \hat{\gamma} y_i}{1 + \hat{\gamma} \hat{\mu}_i} \right) \right].$$

Esto está cerca de la devianza de un GLM Poisson (2.30) cuando $\hat{\gamma}$ esta cerca de 0 (Agresti, 2015).

2.4.4.4. Comparación de Poisson y GLM Binomiales negativos

Para poder hacer la comparación de Poisson y GLM Binomiales negativos que tienen las mismas variables explicativas, y poder determinar si el modelo Binomial negativo da un mejor ajuste, como ya se ha comentado, una comparación informal puede basarse en los valores de AIC. Para una prueba de significancia formal, podemos probar $H_0 : \gamma = 0$, porque el Poisson es el caso limitante de la Binomial negativa a medida que $\gamma \downarrow 0$.

Dado que γ es positivo, $\gamma = 0$ en el límite del espacio del parámetro. Por lo tanto, la estadística de razón de verosimilitud no tiene una distribución chi-cuadrada nula asintótica. Más bien, es una mezcla igual de una distribución de un solo punto en 0 (que ocurre cuando $\hat{\gamma} = 0$) y chi-cuadrada con $g.l. = 1$. El valor p es la mitad que al tratar la estadística como chi-cuadrada con $g.l. = 1$ (Self y Liang, 1987).

2.4.4.5. Modelo Binomial negativo con varianza proporcional a media

Una parametrización Binomial negativa alternativa, resulta de escribir la fórmula de densidad gama con $k\mu$ como el parámetro de forma,

$$f(\lambda; k, \mu) = \frac{k^{k\mu}}{\Gamma(k\mu)} \exp(-k\lambda) \lambda^{k\mu-1} \quad \lambda \geq 0,$$

entonces $E(\lambda) = \mu$ y $\text{var}(\lambda) = \mu/k$. Para esta parametrización, la mezcla gama de distribuciones Poisson produce una distribución Binomial negativa con

$$E(y) = \mu, \quad \text{var}(y) = \mu(1 + k)/k$$

La varianza ahora es lineal en vez de cuadrática en μ . Corresponde a una inflación de la varianza de Poisson, convergiendo a ella a medida que $k \rightarrow \infty$.

Las dos parametrizaciones de la Binomial negativa a veces se denotan por NB1 (lineal) y NB2 (cuadrático). Solo el NB2 cae dentro del marco GLM tradicional, siendo expresable como

una familia de distribuciones de dispersión exponencial, y se usa mucho más comúnmente. A diferencia del modelo NB2, para un modelo NB1, β y k no son parámetros ortogonales, y $\hat{\beta}$ no es un estimador consistente cuando el modelo para la media se mantiene, pero la distribución verdadera no es Binomial negativa (Lee y Nelder, 1996; Cameron y Trivedi, 2013) presentaron el modelo de MV para los modelos NB1 (Agresti, 2015).

2.5. Prueba de sobredispersión

La prueba se define como

$$\text{var}(X) = \mu + cf(\mu)$$

donde $f(\mu)$ es una función de la media μ , además la constante $c < 0$ significa baja dispersión y $c > 0$ significa sobredispersión (Cameron y Trivedi, 1990).

La prueba de hipótesis es

$$\mathcal{H}_0 : c = 0 \quad \text{vs} \quad \mathcal{H}_a : c \neq 0$$

El coeficiente c se estima mediante una regresión de MCO y se prueba con la estadística t (o z) que se distribuye asintóticamente como una normal estándar bajo \mathcal{H}_0 .

3. MATERIALES Y MÉTODOS

El estudio se realizó mediante el programa R Versión 4.0.0, por medio de una simulación Monte Carlo de 10,000 repeticiones para la evaluación y comparación del tamaño de la prueba y la potencia de distintos enfoques de análisis alternativos en experimentos agrícolas.

Se llevaron acabo las pruebas de hipótesis correspondientes para cada modelo mediante un ANOVA para el LM de los modelos Normal y Normal transformado, y ANOVA de GLM Poisson, Quasipoisson y Binomial negativo.

Se consideraron muestras pequeñas con datos que fueran de origen Poisson, es decir, conteos, en donde la media y la varianza son iguales, dicho de otro modo, que los datos no tengan una sobredispersión, en un segundo grupo se analizaron datos con una sobredispersión pequeña, mediana y grande.

Para la obtención de las medias y la varianza se utilizó la herramienta solver, la cual, es un programa del complemento de Microsoft Excel 2019 Versión 16.41 que se usa para encontrar un valor óptimo para una fórmula en una celda, la celda objetivo, para este estudio fue $\delta_{est} - \delta = 0$ que está sujeta a restricciones o limitaciones en los valores de otras celdas de fórmula de una hoja de cálculo del programa a través del Método Simplex; se realizó el cálculo con cuatro y ocho tratamientos, cuatro, seis y ocho repeticiones y los tamaños de efecto pequeño, mediano y grande; esto se hizo para cuando las medias son balanceadas y cuando no lo son, para datos con y sin sobredispersión, las tablas 5.1, 5.2, 5.3 y 5.4 con los resultados se pueden observar en anexos.

Se estimó el parámetro de no centralidad δ mediante la fórmula $\delta_{est} = GL_{Error} \left[\frac{PV}{1-PV} \right]$ presentada por [Murphy y col. \(2014\)](#), además, note que $\delta = n \sum \frac{(\mu_t - \mu)^2}{\sigma^2}$ donde n es el número de repeticiones por tratamiento, μ_t es la media de los t tratamientos, μ la media general y σ^2 la varianza debida al error de muestreo.

Se utilizó la prueba de sobredispersión del paquete AER de R para verificar si los datos,

a través del modelo GLM Poisson, contaban con sobredispersión, o no.

En cada caso, se obtuvieron los datos a través de todas las combinaciones de los factores que fueron estudiados y se registro el número de rechazos de las pruebas de hipótesis obtenidos de las 10,000 repeticiones. El nivel de significancia α se estableció en 0.05 para probar el juego de hipótesis

$$\mathcal{H}_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0 \quad \text{vs} \quad \mathcal{H}_a : \text{al menos un } \tau_i \neq 0.$$

La potencia de la prueba fue estimada mediante la proporción de rechazos cuando el tamaño del efecto fue diferente de cero. El tamaño de prueba se estimó como la proporción de rechazos cuando el tamaño del efecto fue cero, esto mismo utilizado por [Declercq y col. \(2019\)](#). Para la prueba de sobredispersión se realizó el mismo procedimiento, que para la potencia.

3.1. Factores de estudio

3.1.1. Modelos

Los modelos utilizados fueron LM Normal y Normal transformado, y GLM Poisson, Quasi-poisson y Binomial negativo. Éstos modelos representan, en su mayoría, a los modelos utilizados en agronomía para el análisis de variables respuesta de conteos, por lo que es importante el uso adecuado de ellos, apegado a la realidad y que cuenten con la suficiente potencia estadística.

3.1.2. Número de repeticiones

Los experimentos agrícolas en general utilizan tamaño de muestra muy pequeños, ya que en la agronomía es difícil el contemplar tener muchas repeticiones, por lo que en este estudio se considerarán cuatro, seis y ocho repeticiones, además de este número, se utilizaron cincuenta repeticiones para observar el comportamiento de las muestras con un buen número de repeticiones.

3.1.3. Número de tratamientos

La experiencia, ha mostrado que experimentos muy grandes disminuyen su probabilidad de éxito, en agricultura esto se ve acentuado debido a que los campos experimentales presentan variabilidad por su naturaleza, y experimentos grandes incrementan el riesgo de tener un

buen manejo de los factores incontrolables que se presentan durante el experimento, por lo que para este estudio, se decidió utilizar cuatro tratamientos como un experimento pequeño y ocho tratamientos como un experimento grande, ya que uno de los factores que afecta a la potencia, como se puede ver en 2.12, es el número de tratamientos.

3.1.4. Tamaño del efecto

El tamaño del efecto es uno de los factores que más afecta a la potencia en un trabajo de investigación. En muchas áreas de estudio, existe un cuerpo sustancial de teoría e investigación empírica que proporcionará una estimación bien fundamentada del tamaño del efecto, [Murphy y col. \(2014\)](#) mencionan tres diferentes tamaños de efecto, 0.01, 0.1 y 0.25, pequeño, mediano y grande, respectivamente, los cuales fueron utilizados en este estudio.

3.1.5. Distribución entre medias

Las medias se consideraron de dos formas, balanceadas y desbalanceadas, para el caso de medias balanceadas se consideraron cuando son cuatro tratamientos, dos medias iguales, y para cuando son ocho, se tomaron cuatro iguales; para el caso de medias desbalanceadas, cuando son cuatro tratamientos se utilizaron tres medias iguales y una diferente, para el caso de ocho tratamientos, seis medias iguales y dos diferentes.

3.1.6. Sobredispersión

En agronomía se pueden encontrar muchas variables de respuesta que son conteos, las cuales regularmente, son analizadas mediante modelos Poisson, donde uno de sus supuestos es que la media es igual a la varianza, pero esto en la práctica, es muy difícil que ocurra, de aquí que los datos generen una sobredispersión, teniendo varianzas que se alejan de su media.

Se analizaron dos casos de datos con variable respuesta Poisson, cuando no existe una sobredispersión y cuando si existe, el parámetro de forma de sobredispersión utilizado como base fue tomada del ejemplo de cangrejos herradura de [Agresti \(2015\)](#), por lo que los valores utilizados se pueden observar en la tabla 3.1.

Tabla 3.1. Parámetros de forma de la sobredispersión (k) utilizados.

Grande	Mediana	Pequeña
20	40	100

Tomando en cuenta que, en la familia Binomial negativa de dos parámetros, donde el parámetro de forma $k > 0$, entonces $\gamma = 1/k$ y que y tiene $E(y) = \mu$ y $\text{var}(y) = \mu + \gamma\mu^2$. El índice $\gamma > 0$ es un tipo de parámetro de dispersión.

Las tablas [5.1](#), [5.2](#), [5.3](#) y [5.4](#) en anexos, presentan las medias utilizadas para los diferentes casos, sin y con sobredispersión para cuatro y ocho tratamientos, con medias balanceadas y desbalanceadas.

4. RESULTADOS Y DISCUSIÓN

Tomando en cuenta que las variables resultantes son binarias, debido a que en una prueba de hipótesis se rechaza o no se rechaza, se utilizó un modelo de regresión logística para estimar los efectos de los factores estudiados en la probabilidad de rechazo, y así ayudar en la interpretación de los datos y facilitar el análisis; los resultados están divididos en dos partes, una sección para el tamaño de la prueba, y otra sección para la potencia.

Al momento de hacer la evaluación y discusión para los diferentes modelos, se dividió en dos casos, cuando las medias son balanceadas y cuando no lo son, los ejemplos utilizados fueron para cuando los tratamientos son cuatro y ocho tratamientos, tres tamaños de efecto que son pequeño, mediano y grande con cuatro, seis, ocho y cincuenta repeticiones.

4.1. Tamaño de la prueba

4.1.1. Sin sobredispersión (datos Poisson)

Para las simulaciones cuando los datos son Poisson, es decir, que no tienen sobredispersión, en todos los modelos el α nominal se mantiene por debajo de 0.05, excepto el modelo Quasipoisson.

Como se puede observar en la figura 4.1, el modelo Quasipoisson en todos los casos, no conserva el valor del α nominal, a excepción del caso con cincuenta repeticiones, se puede ver que a medida que las repeticiones aumentan, el valor disminuye hasta llegar a 0.05; El modelo Binomial negativo se mantiene por debajo del valor de $\alpha = 0.05$ nominal en todas las repeticiones. [Declercq y col. \(2019\)](#) encontraron para muestras pequeñas que, la tasa de error tipo I del LM se encuentra más bajo control que cuando se usa GLM, similarmente, [Ives \(2015\)](#) demostró a través de simulación que los LM, en comparación con los GLM, tienen un control de error de tipo I más confiable y pierden poca potencia, por lo que brindan una alternativa sólida.

Para el modelo Normal, Normal transformado y Poisson en todas las repeticiones se mantiene su valor del α nominal en 0.05, esto reafirma lo obtenido por Lewis y col. (2001) , en donde encontraron que para el modelo Poisson mediante simulación Monte Carlo para muestras pequeñas, la cobertura observada está muy cerca de la afirmación teórica asintótica del 95 %.

Se observa en la figura 4.1 que el tamaño de la prueba se conserva mejor a medida que aumentamos en las repeticiones.

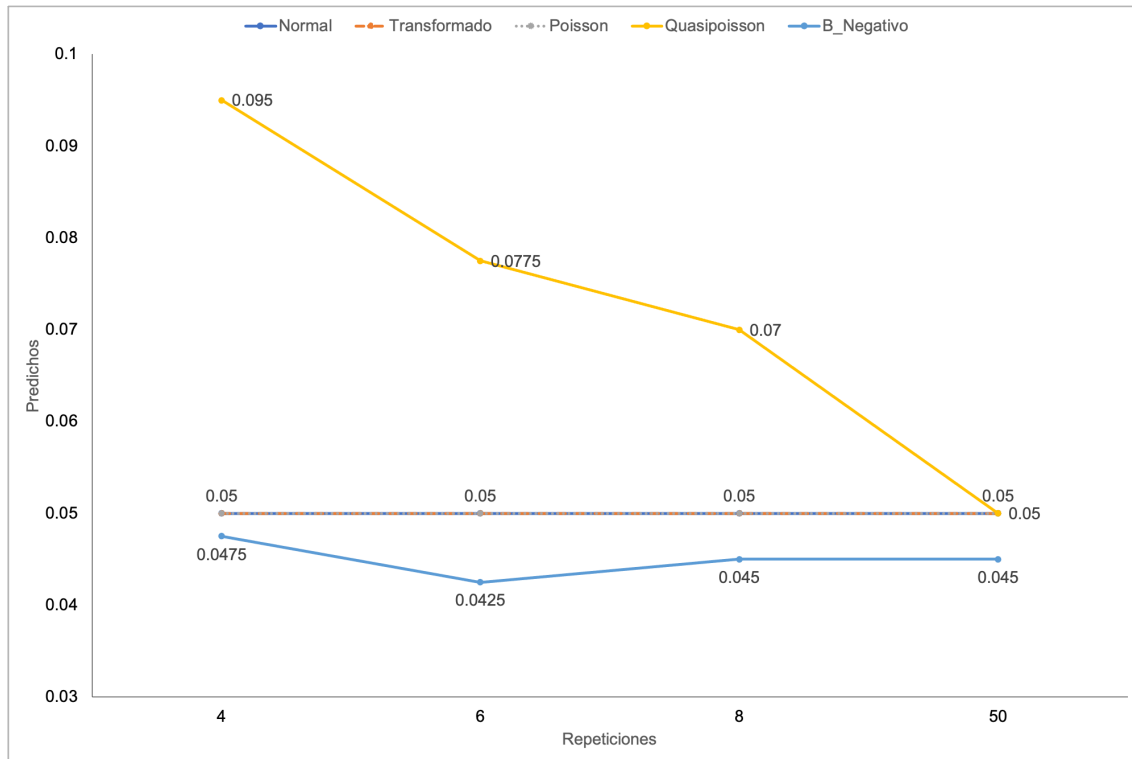


Figura 4.1. Comparación del tamaño de la prueba promedio de los modelos para datos Poisson a través de diferentes repeticiones.

Se encontró interacción entre el número de tratamientos y repeticiones en cuanto al tamaño de la prueba. La figura 4.2 muestra la forma en que se presenta esta interacción.

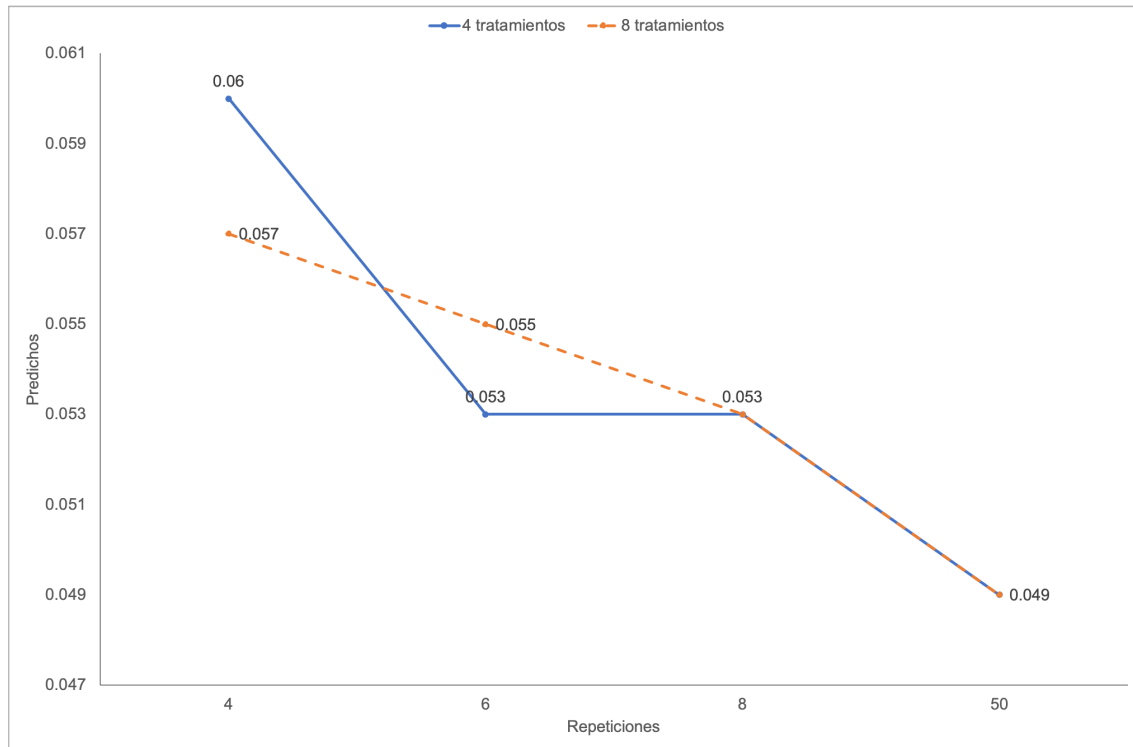


Figura 4.2. Comparación del tamaño de la prueba promedio de los tratamientos para datos Poisson a través de diferentes repeticiones.

Cuando utilizamos un mayor desbalance entre medias, véase figura 4.3 el α nominal se conserva mejor y se puede apreciar nuevamente que, a medida que aumentan las repeticiones se conserva el valor de α nominal.

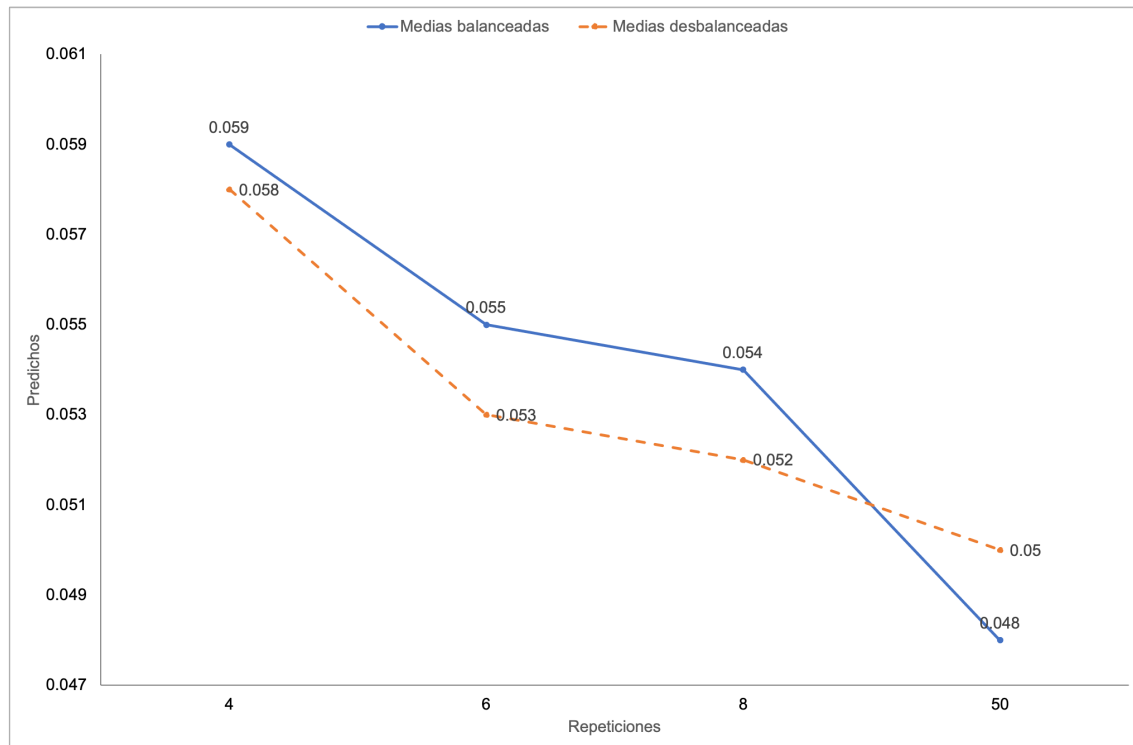


Figura 4.3. Comparación del tamaño de la prueba promedio de las distintas medias para datos Poisson a través de diferentes repeticiones.

4.1.2. Sobredispersión grande

Cuando tenemos valores simulados con una sobredispersión grande tenemos que el modelo Poisson, Binomial negativo y Quasipoisson no conservan el tamaño de prueba, solamente el modelo Normal y Normal transformado como se puede ver en la figura 4.4. Young y col. (1999) consideran los ANOVA aplicados para datos de conteos y muestran mediante simulación Monte Carlo que, los LM tienen mejores tasas de error de tipo I que los GLM comparadas con el $\alpha = 0.01, 0.05$ y 0.10 nominal para identificar los efectos del tratamiento cuando existe una sobredispersión grande, un número grande de tratamientos y un número pequeño de repeticiones por tratamiento. Note que cuando pasamos de cuatro a ocho tratamientos en los modelos Normal, Normal transformado y Quasipoisson se conserva el mismo valor del α nominal, a diferencia de los modelos Binomial negativo y Poisson que aumentan. Se puede ver que, el modelo que conserva menos el valor del α nominal es el Poisson, seguido del Binomial negativo y por último el Quasipoisson.

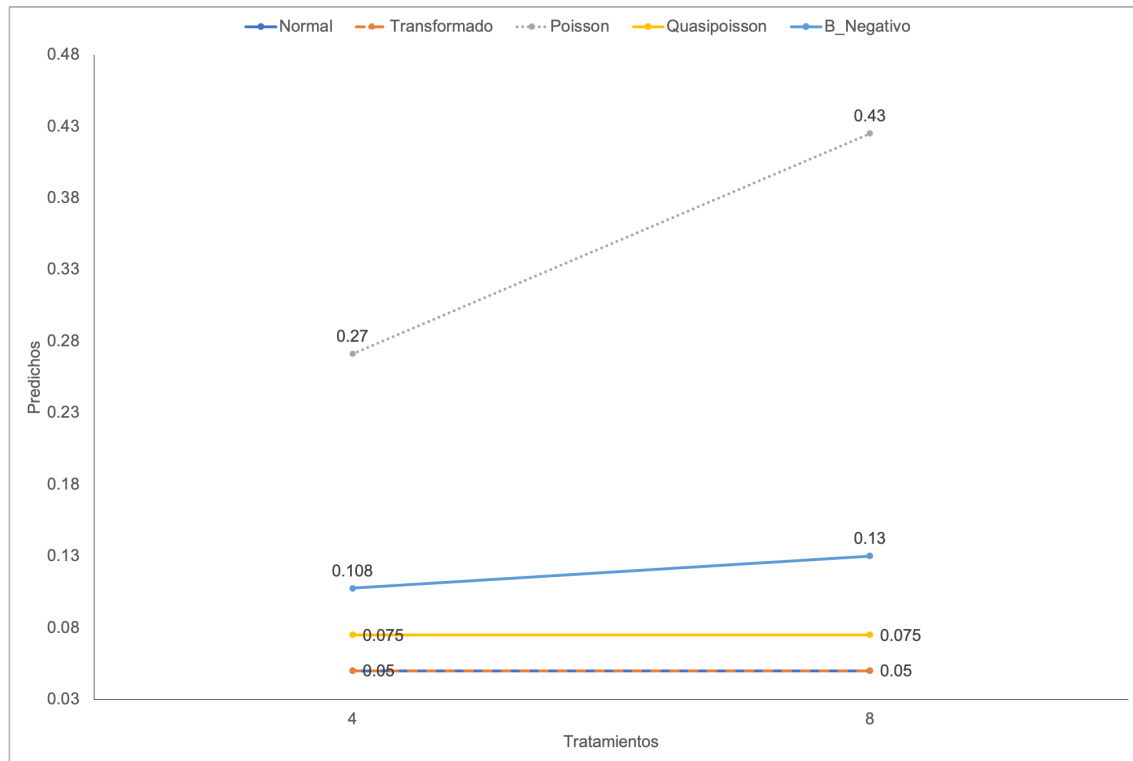


Figura 4.4. Comparación del tamaño de la prueba promedio de los modelos para datos con sobredispersión grande a través de diferentes tratamientos.

Se puede apreciar en la figura 4.5 que, el modelo Normal y Normal transformado mantienen su valor de $\alpha = .05$ nominal en todos los casos. Los modelos Quasipoisson, Binomial negativo solamente lo conservan para cincuenta repeticiones, es decir, éstos dos modelos tienen un comportamiento adecuado en muestras grandes. A medida que aumentamos en las repeticiones a cincuenta, todos los modelos excepto el Poisson tienden a acercarse al valor del α nominal.

En todos los casos, el valor del tamaño de las pruebas del modelo Poisson, se encuentra muy alejado del tamaño nominal. Warton y col. (2016) nos dicen que los LM serán robustos a la sobredispersión, a diferencia de los GLM, pero se debe de tomar en cuenta que, los LM pueden tener una pérdida sustancial de potencia en comparación con los GLM.

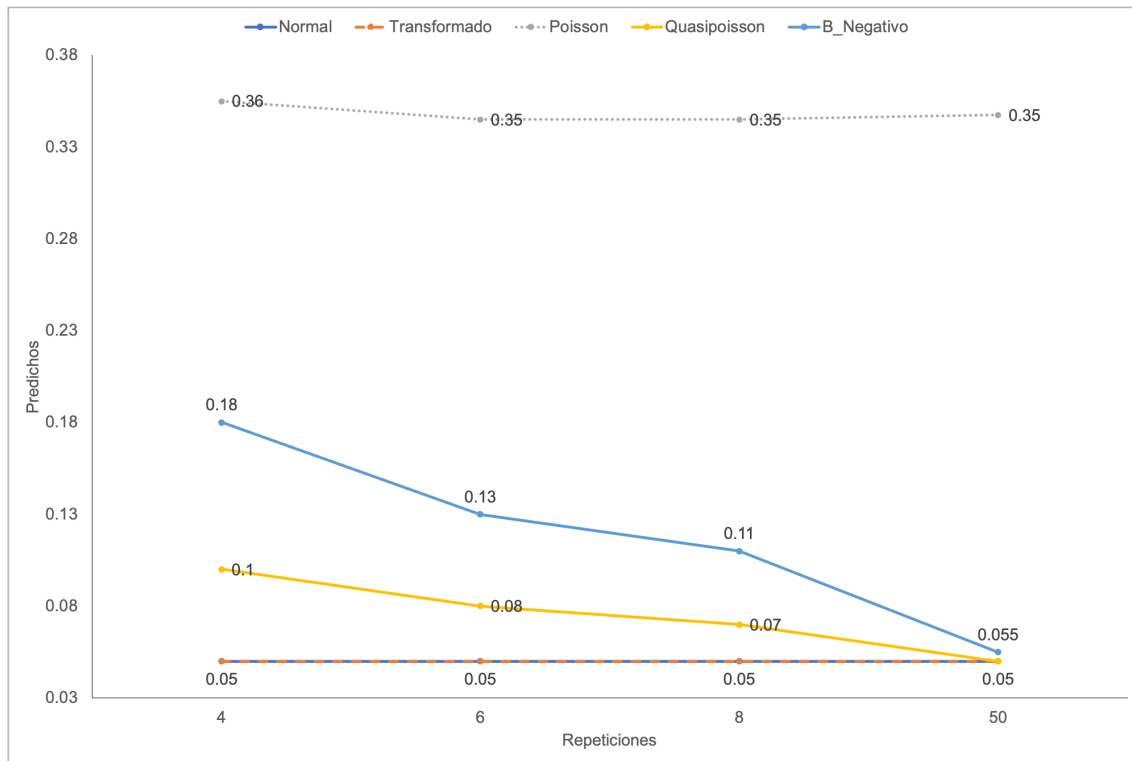


Figura 4.5. Comparación del tamaño de la prueba promedio de los modelos para datos con sobredispersión grande a través de diferentes repeticiones.

Se advierte en la figura 4.6 que, tanto para cuatro y ocho tratamientos, el tamaño de la prueba para cuando tenemos menos repeticiones es mayor, por lo contrario, cuando tenemos más repeticiones el tamaño de la prueba tiende a ser menor.

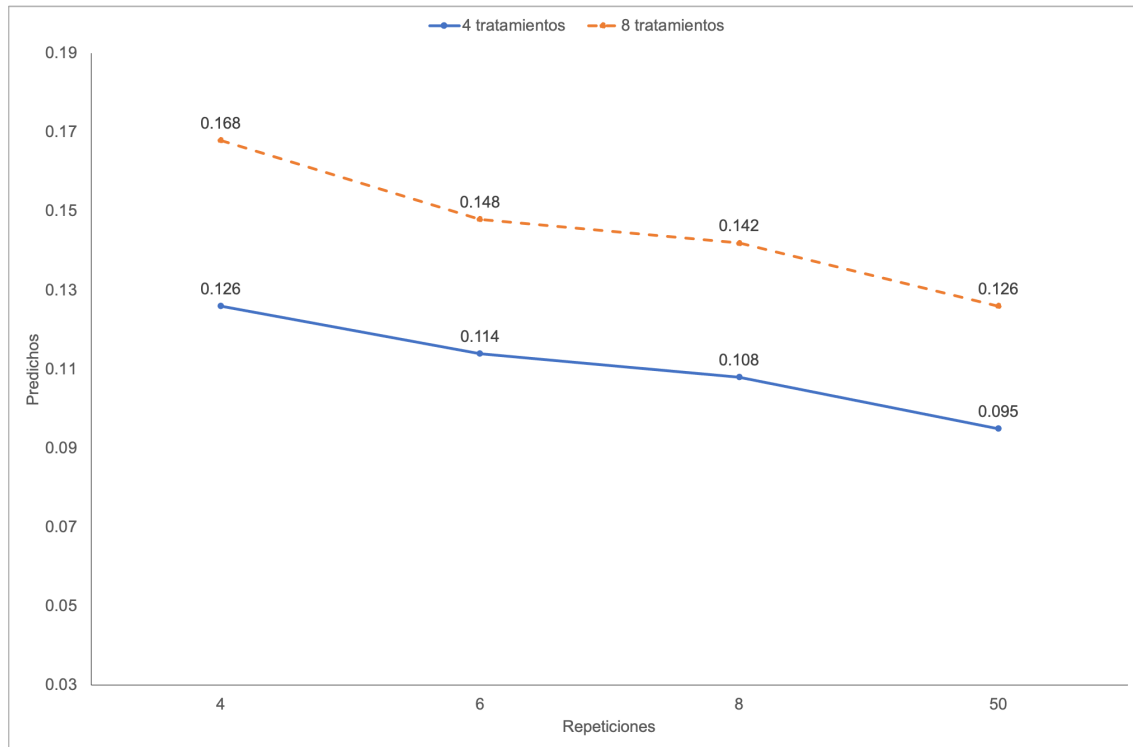


Figura 4.6. Comparación del tamaño de la prueba promedio de los tratamientos para datos con sobredispersión grande a través de diferentes repeticiones.

4.1.3. Sobredispersión mediana y pequeña

Cuando utilizamos una sobredispersión mediana o pequeña observamos que ocurre lo mismo que cuando utilizamos una sobredispersión grande, es decir, los modelos Normal y Normal transformado mantienen su α nominal en 0.05 seguidos del Quasipoisson, Binomial negativo y por último el Poisson. En el modelo Quasipoisson vemos que el valor de α nominal es mayor cuando tenemos una sobredispersión media que una grande; Para los modelos Binomial negativo y Poisson el valor del α nominal es menor en sobredispersión mediana y pequeña que cuando tenemos una sobredispersión grande para ocho tratamientos, esto se puede percibir en la figura 4.7, para el caso de cuatro tratamientos el comportamiento es similar.

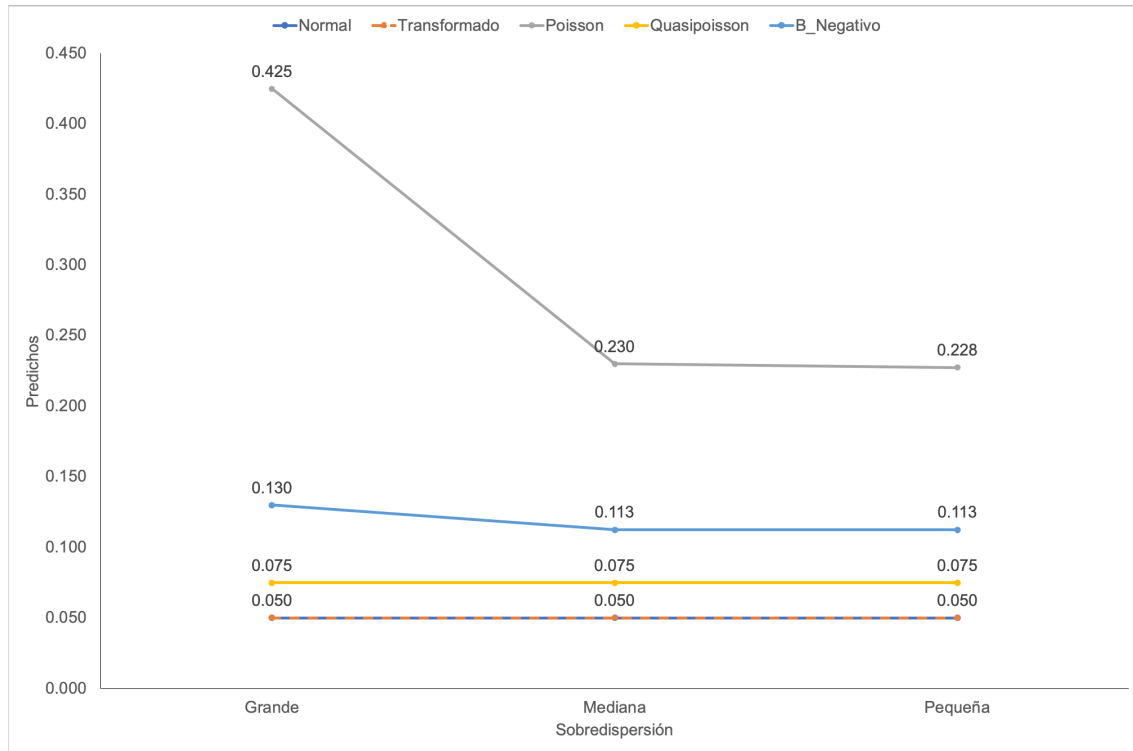


Figura 4.7. Comparación del tamaño de la prueba promedio de los modelos a través de la sobredispersión, caso de ocho tratamientos.

El caso en la figura 4.8 es para cuando tenemos cuatro repeticiones, la comparación se hace para las tres sobredispersiones utilizadas, en donde el valor de α nominal se mantiene para los modelos Normal y Normal transformado en los tres casos, el modelo Quasipoisson disminuye un poco para cuando se tiene una sobredispersión pequeña; Para el modelo Binomial negativo y Poisson se puede ver que tiende a bajar a medida que la sobredispersión disminuye, donde en el modelo Poisson se puede apreciar una pendiente negativa mayor, cuando se pasa de una sobredispersión grande a una mediana.

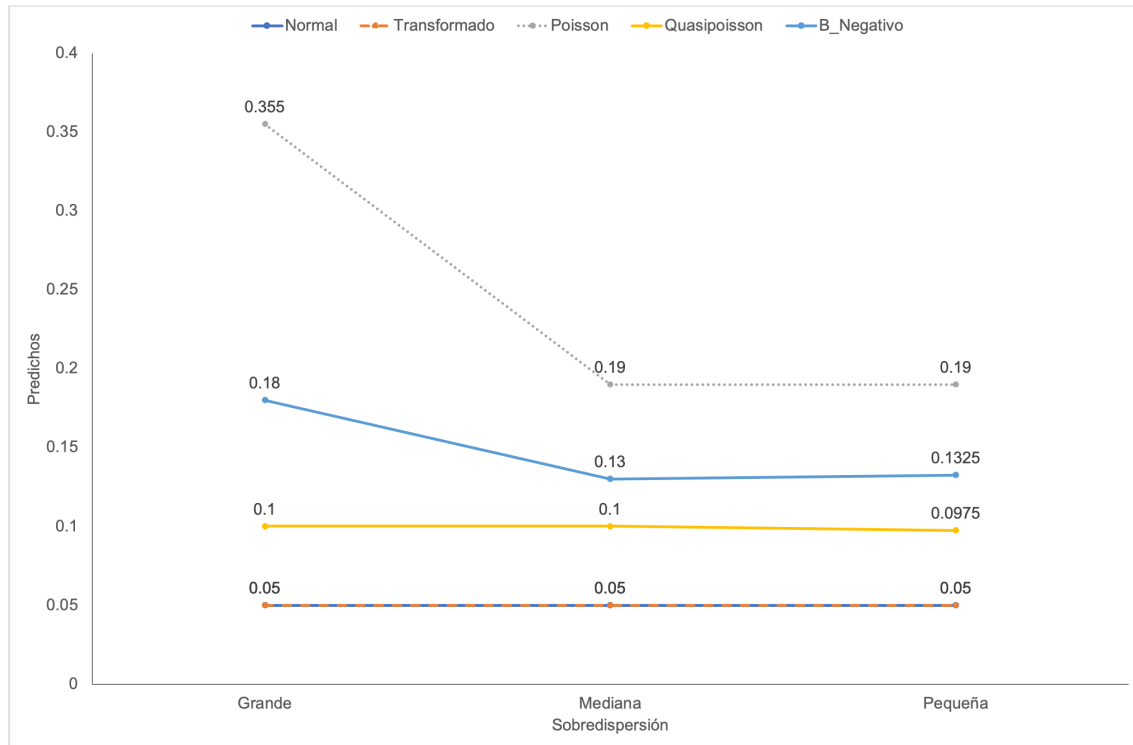


Figura 4.8. Comparación del tamaño de la prueba promedio de los modelos a través de la sobredispersión, caso de cuatro repeticiones.

Para tener un análisis más cercano a la realidad que se quería simular de las sobredispersiones, como se vio en los dos análisis anteriores se hizo la comparación cuando se tienen cuatro repeticiones, esto es debido a que es la mayor frecuencia en los análisis de agronomía, por lo que al momento de comparar entre cuatro y ocho tratamientos, podemos ver en la figura 4.9 que el tamaño de la prueba se conserva mejor para cuatro que para ocho, y es mejor para cuando tenemos menos sobredispersión.

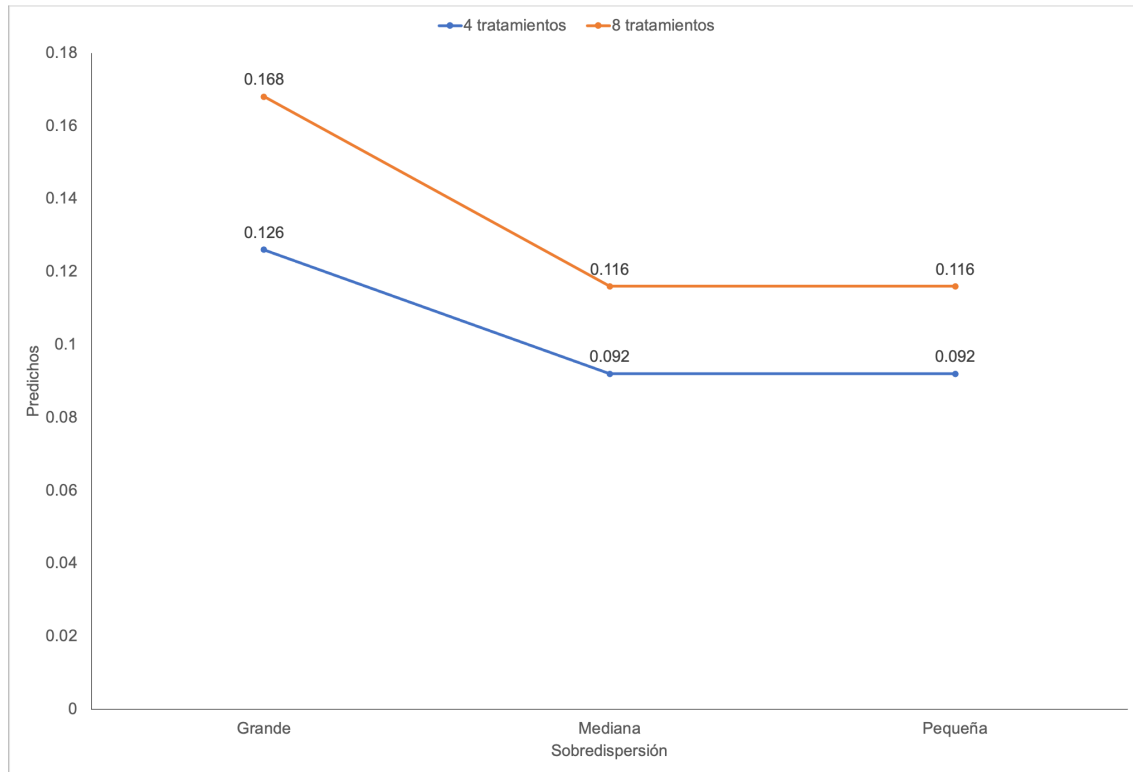


Figura 4.9. Comparación del tamaño de la prueba promedio de los diferentes tratamientos a través de la sobredispersión, caso de cuatro repeticiones.

Note que, los modelos Normal y Normal transformado mantienen el valor de α nominal en los tres tipos de sobredispersión utilizados, el modelo Quasipoisson y Binomial negativo solamente al momento de utiliza cincuenta repeticiones. En el modelo Binomial negativo podemos decir que conforme la sobredispersión va disminuyendo se acerca más al valor de $\alpha = 0.05$ nominal, pero se mantiene arriba de 0.1. En el modelo Poisson se tienen los cambios más significativos en este modelo al pasar de una sobredispersión grande a una mediana, pero en ninguno de los casos planteados mantiene el valor del α nominal, siendo este modelo el que tiene los valores más altos, este tipo de resultados también fueron encontrados por (Gardner y col., 1995) al comparar los modelos Poisson y Binomial negativo para analizar datos de conteos mediante regresión lineal para datos psicológicos.

4.2. Potencia

4.2.1. Sin sobredispersión

Al evaluar las potencias se dividió de la misma manera que para el tamaño de la prueba, agregando una partición más, cuando utilizamos cuatro y ocho tratamientos.

Cuando utilizamos cuatro tratamientos podemos ver en la figura 4.10 que, el modelo Quasi-poisson es el que mayor potencia tiene, pero debido a que no conserva el valor del α nominal, se omitirá en las siguientes figuras; Cuando tenemos un tamaño de efecto pequeño no existe diferencia entre los modelos, esto lo comenta también Jiménez (2019) al hacer la comparación entre los modelos Poisson y Normal; a medida que el efecto aumenta la potencia de los modelos tiende a separarse; el modelo que mayor potencia presenta es el Poisson seguido del Binomial negativo y en un tercer lugar se encuentran el Normal y Normal transformado en las mismas circunstancias, esto se puede observar en los tamaños de efecto mediano y grande. Al hacer una comparación en las potencias de modelos LM y GLM, (Hrdličková, 2006; Declercq y col., 2019) también encontraron que el GLM tiene una mayor potencia que un LM cuando se tienen muestras pequeñas.

El modelo Poisson y el modelo Binomial negativo se comportan de la misma manera para un tamaño del efecto pequeño, a medida que aumenta el efecto éstos se separan, siendo el modelo Poisson el que tiene mayor potencia. Para los modelos Normal y Normal transformado no existe una diferencia, es decir, la potencia es ascendente a medida que el tamaño del efecto aumenta, pero siempre por debajo del Poisson y Binomial negativo. Note que, la potencia no es mayor a 0.669 cuando se utilizan cuatro tratamientos y 0.801 para ocho tratamientos como se muestra en la figura 4.11, Cohen (1988, 1992) y Brown (2015) comentan que, un nivel deseado de potencia estadística es 0.8 y un mínimo de 0.5, es decir, un 50 % de detectar un efecto existente.

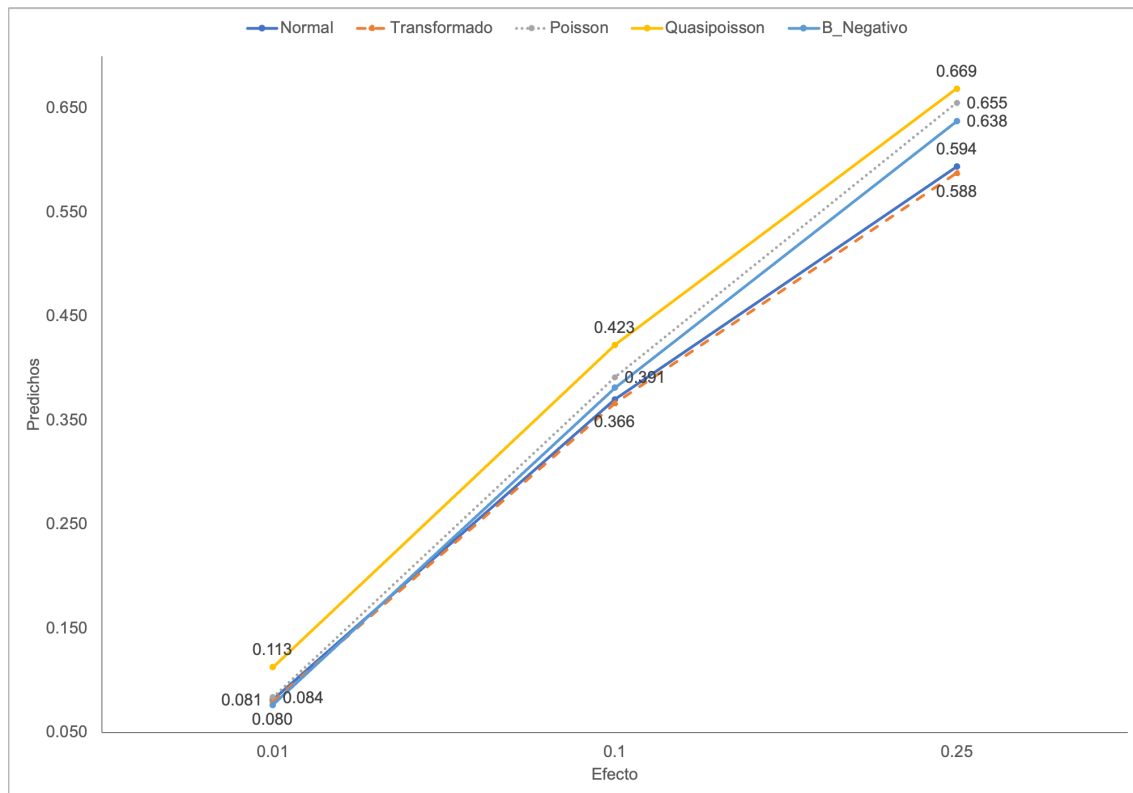


Figura 4.10. Comparación de la potencia de la prueba de modelos para datos Poisson a través del tamaño del efecto.

Para ver la diferencia que existe entre utilizar cuatro y ocho tratamientos y su implicación, se tomo el tamaño de efecto grande y se compararon como se puede apreciar en la figura 4.11, note que en todos los modelos existe un aumento en la potencia a medida que aumentamos los tratamientos, siendo el cambio menos significativo en el modelo Normal, y el modelo con mayor diferencia entre tratamientos, el Binomial negativo.

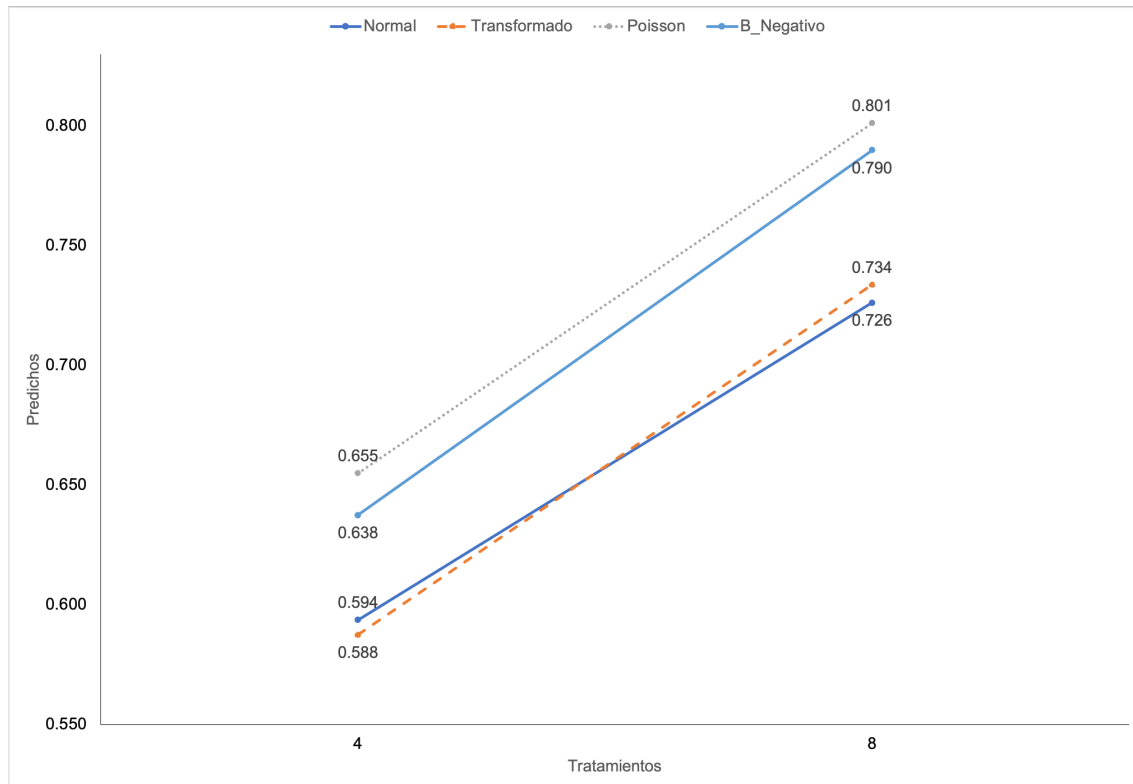


Figura 4.11. Comparación de la potencia de la prueba promedio de modelos para datos Poisson a través de los tratamientos, caso del tamaño de efecto grande.

En el caso de no sobredispersión, existe una interacción triple, por lo que se generaron dos gráficas, como podemos notar en la figura 4.12a cuando tenemos los casos balanceado y desbalanceado, podemos ver que, en el tamaño de efecto pequeño en las repeticiones cuatro, seis y ocho, el valor de la potencia es muy similar, dicho de otro modo, las diferencias entre cuatro, seis y ocho repeticiones no son tan marcadas a cuando tenemos cincuenta repeticiones, la potencia es mayor a las anteriores, siendo aún más marcada en el caso de medias desbalanceadas. En los tamaños de efecto mediano y grande, a medida que las repeticiones aumentan la potencia aumenta, entre más repeticiones y mayor tamaño de efecto, las diferencias en las potencias son mayores de manera general, esto lo comenta también Lewis y col. (2001). Podemos observar que, en el caso balanceado, las diferencias en el efecto mediano son menores entre las repeticiones que en el caso desbalanceado. El hecho de que haya interacción es porque la magnitud con que incrementa la potencia no es la misma para las repeticiones conforme el efecto es mayor dependiendo el caso, para el caso balanceado los incrementos son mayores.

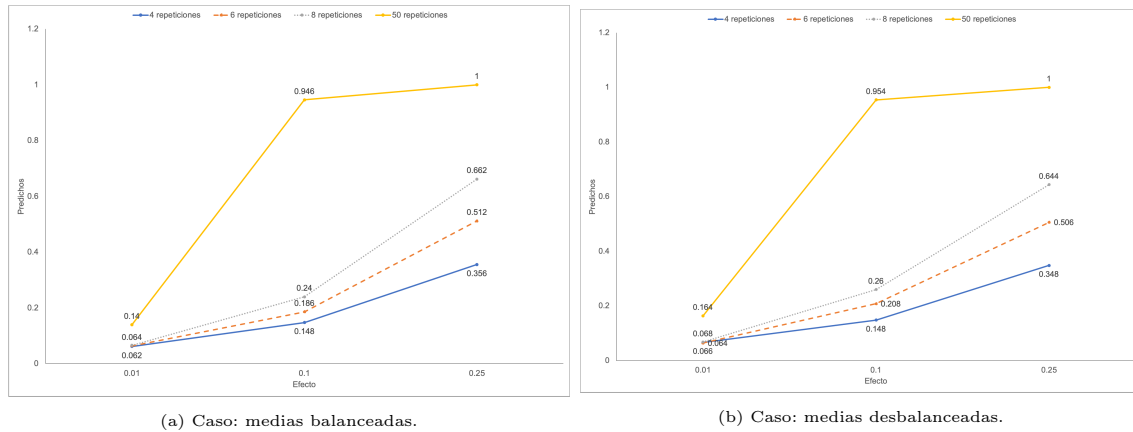


Figura 4.12. Comparación de la potencia de la prueba promedio de las diferentes repeticiones para datos Poisson a través del tamaño del efecto.

Cuando utilizamos ocho tratamientos la potencia tiende a ascender con mayor rapidez conforme el efecto aumenta a comparación con el de cuatro, siendo en el caso de ocho repeticiones la diferencia mayor entre potencias; En todas las repeticiones y todos los tamaños de efecto, la potencia es mayor para el caso de ocho tratamientos. Tomando ahora el tamaño de efecto mediano y el caso balanceado, podemos advertir, como lo muestra la figura 4.13, lo anterior mencionado.

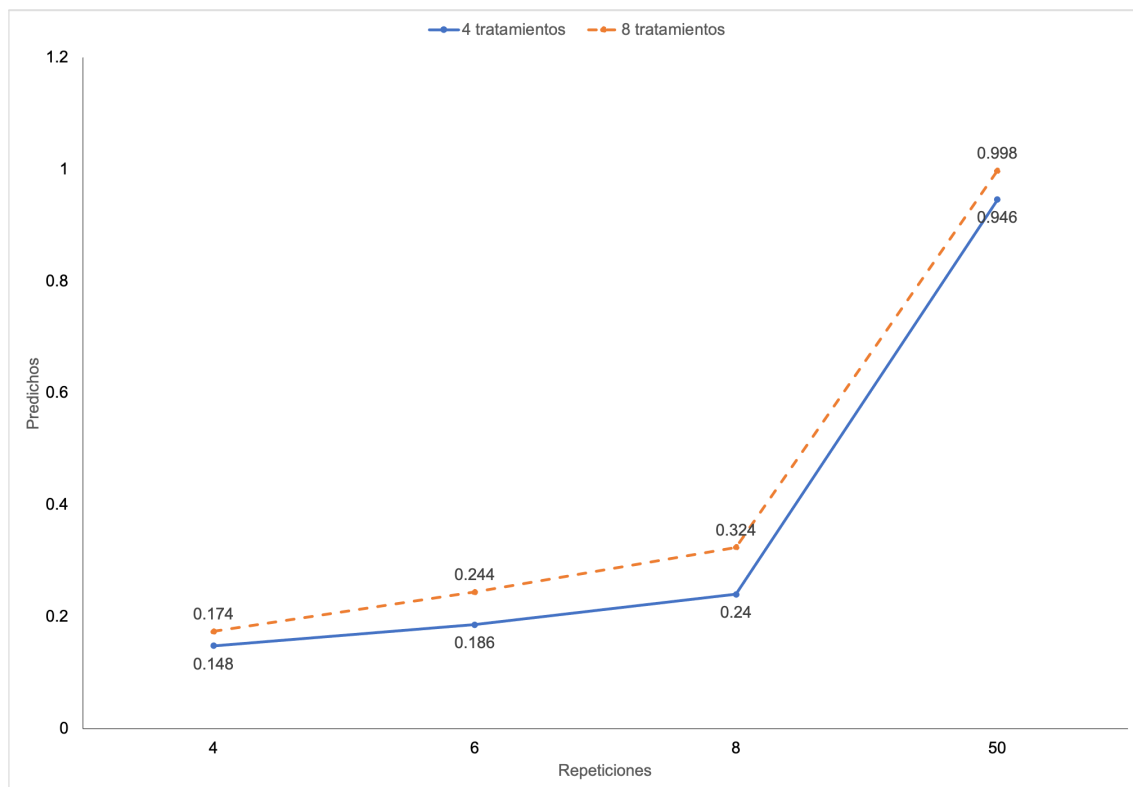


Figura 4.13. Comparación de la potencia de la prueba de los tratamientos para datos Poisson a través de las diferentes repeticiones, caso de medias balanceadas y tamaño del efecto mediano.

Todos los modelos se comportan de manera ascendente conforme se aumentan el número de repeticiones, a medida que el tamaño de la muestra aumenta, la potencia de la prueba aumenta, esto lo comenta también (Montgomery, 2017), siendo el modelo Poisson el que más aumenta en el cambio de cuatro a seis repeticiones. Cuando tenemos cuatro repeticiones las diferencias se ven más marcadas en la potencia, donde el modelo Normal y Normal transformado tienen la menor potencia, seguidos del Binomial negativo y por último el modelo Poisson. En este caso podemos ver que existe interacción, ya que la potencia responde menos en los modelos Normal, Normal transformado y Binomial negativo que en el modelo Poisson, como nos muestra la figura 4.14.

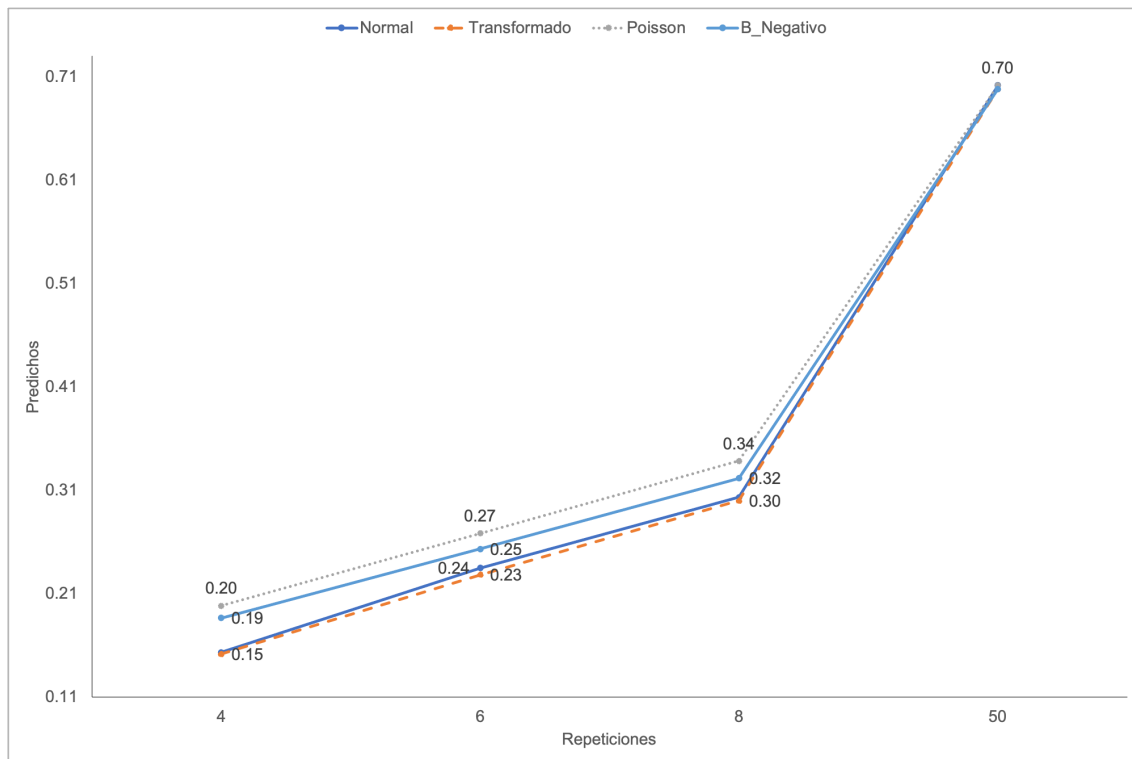


Figura 4.14. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de las diferentes repeticiones.

Si tomamos el caso de cuatro repeticiones y hacemos la comparación de los modelos entre cuatro y ocho tratamientos, se observa en la figura 4.15 que las potencias se comportan de la misma manera ascendente, para el caso del modelo Normal y Normal transformado existe una pequeña diferencia entre los modelos, tanto para cuatro, como para ocho tratamientos; para los modelos Binomial negativo y Poisson, la diferencia entre las potencias para cuatro y ocho tratamientos tiende a ser la misma.

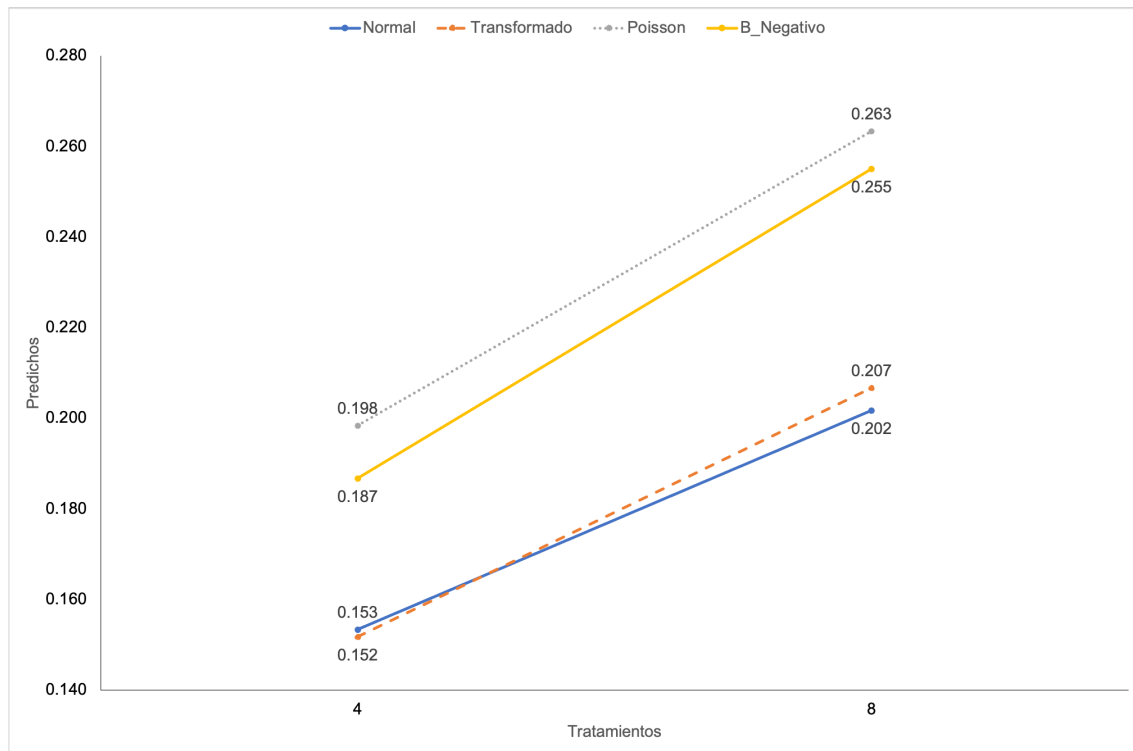


Figura 4.15. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de los tratamientos, caso de cuatro repeticiones.

Cuando se analizan las repeticiones y las medias balanceadas y desbalanceadas, podemos ver que, para cuatro y ocho repeticiones la potencia es la misma para los dos casos, se puede apreciar en la figura 4.16 cuando utilizamos seis y cincuenta repeticiones que, la potencia es mayor para el caso desbalanceado.

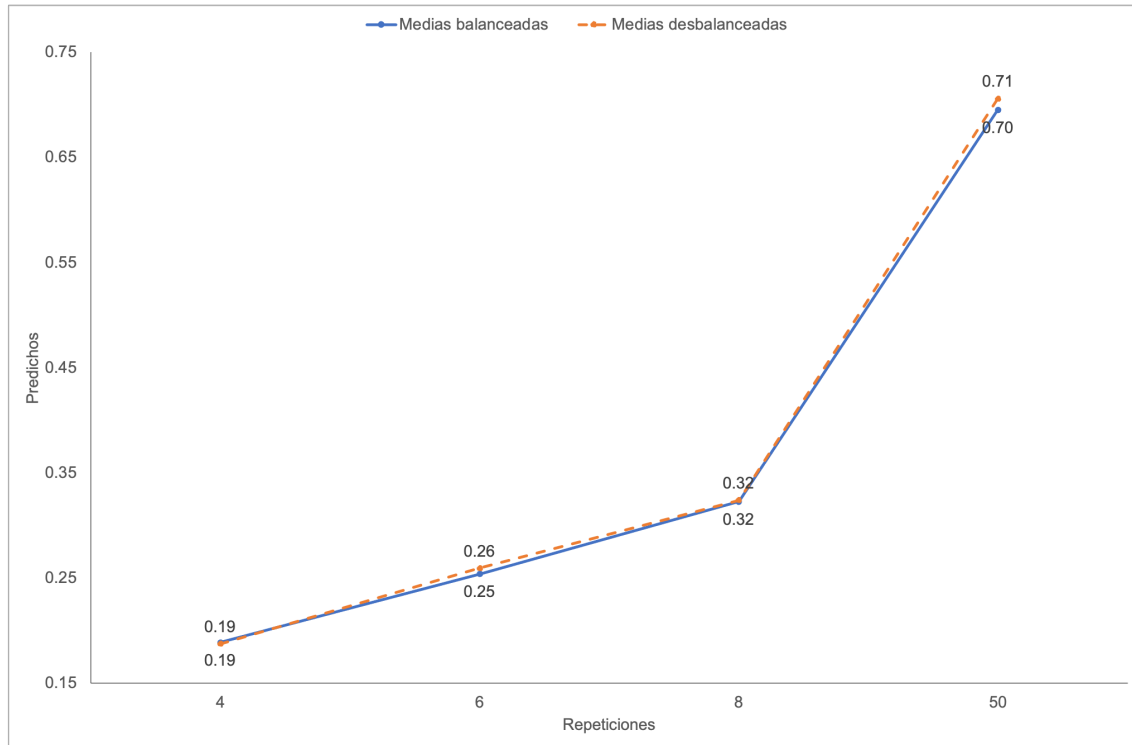


Figura 4.16. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través de las diferentes repeticiones, caso de cuatro tratamientos.

A diferencia de cuando se utilizan cuatro tratamientos, al utilizar ocho, existen diferencias en las potencias en todas las repeticiones, siendo el desbalanceado el que tiene mayor potencia; Si comparamos las potencias entre cuatro y ocho tratamientos se observa en la figura 4.17 que, en todas las repeticiones la potencia es mayor cuando utilizamos ocho tratamientos.

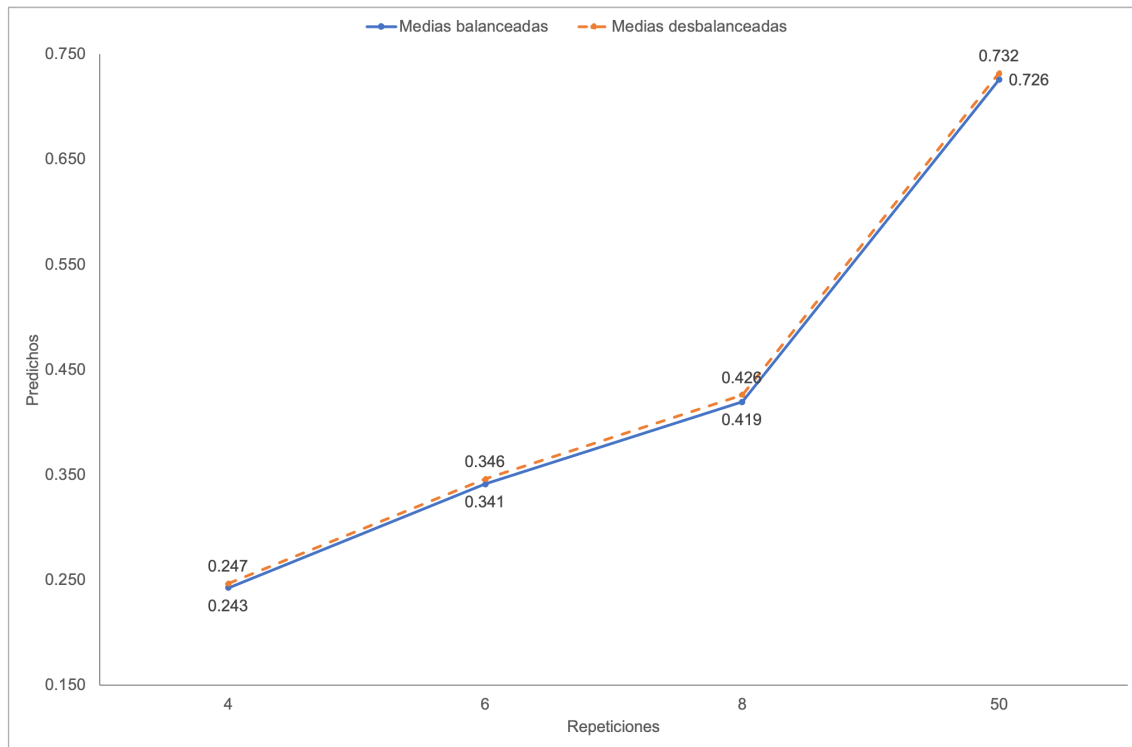


Figura 4.17. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través de las diferentes repeticiones, caso de ocho tratamientos.

Note en la gráfica 4.18 que, cuando tenemos efecto pequeño y mediano la potencia es mayor para el caso desbalanceado y para el tamaño de efecto grande, la potencia mayor es cuando tenemos datos balanceados.

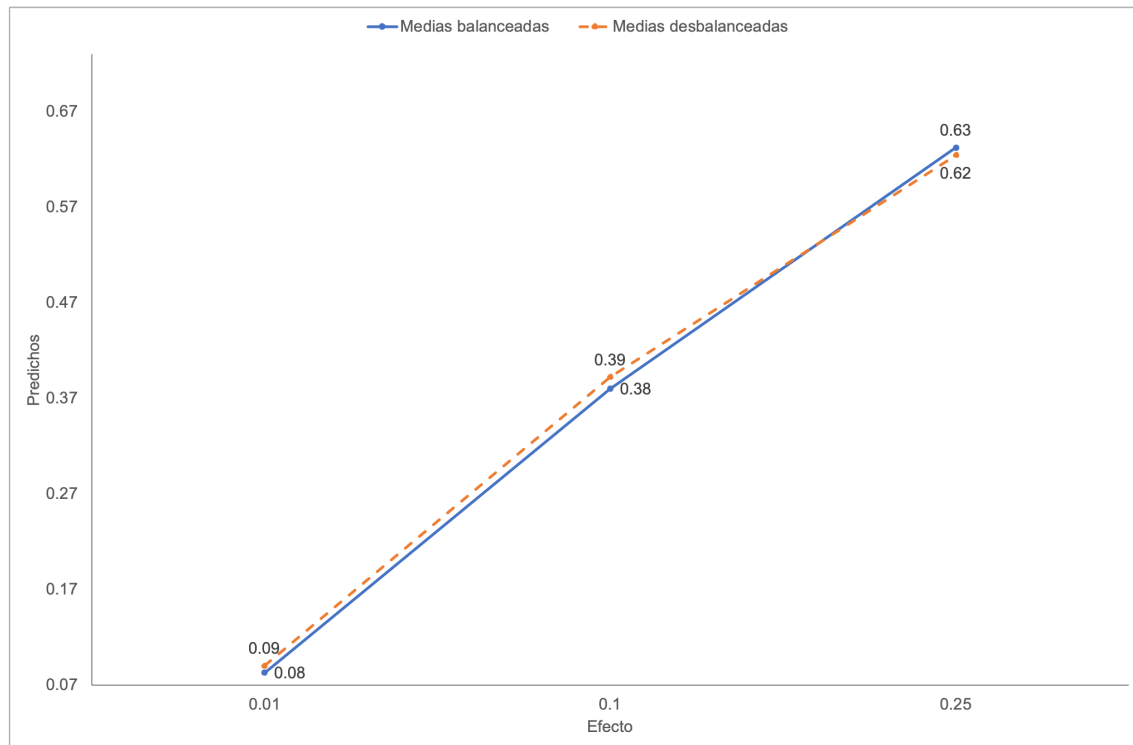


Figura 4.18. Comparación de la potencia de la prueba de las diferentes medias para datos Poisson a través del tamaño del efecto, caso de cuatro tratamientos.

Para cuando utilizamos cuatro tratamientos no se encontró interacción entre los casos y los modelos, más que para ocho, teniendo que el modelo Normal es el que tiene menor potencia, seguido del Normal transformado, Binomial negativo, y al final el Poisson con la mayor potencia en ambos casos; En todos los modelos se nota que existe una potencia mayor para el caso de medias desbalanceadas que para medias balanceadas, se observa en la figura 4.19 que para el modelo Normal transformado el incremento es mayor entre casos.

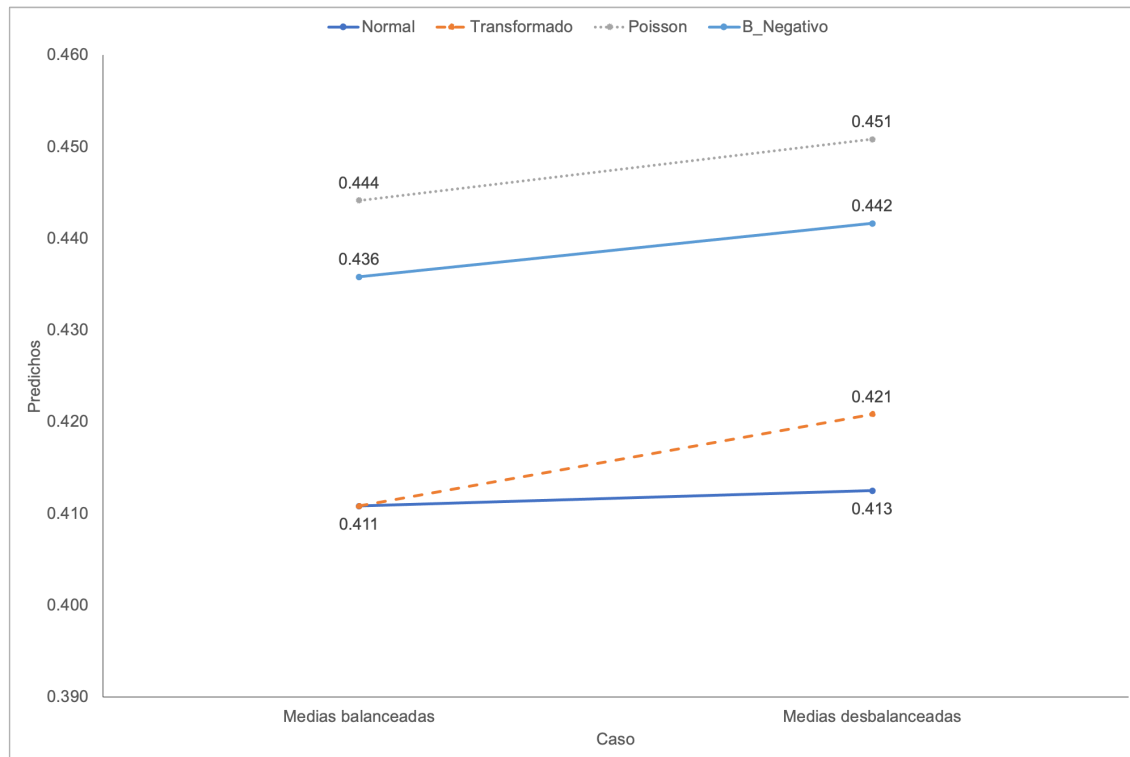


Figura 4.19. Comparación de la potencia de la prueba de los modelos para datos Poisson a través de las diferentes medias, caso de ocho tratamientos.

4.2.2. Sobredispersión grande

En los datos de la sobredispersión grande, con respecto a su potencia, observamos en la figura 4.20a, la cuál nos presenta la interacción entre el tamaño del efecto y los modelos en el caso balanceado que, a medida que el tamaño del efecto aumenta, aumenta la potencia, es decir, los casos balanceados con un tamaño de efecto pequeño tendrán una potencia entre 0.083 y 0.338 en donde los modelos Normal y Normal transformado son los que poseen la menor, y el modelo Poisson la mayor, recordando que al momento de discutir el tamaño de la prueba con sobredispersión, el modelo Poisson es el que menos conservaba su valor, por lo que en las siguientes discusiones, no se tomará en cuenta, esto es consistente con lo que comenta Hilbe (2017), cuando haya una sobredispersión, no se debe utilizar un modelo Poisson. Además, Jiménez (2019) asegura que en un caso de heteroscedasticidad alta, el intervalo de confianza del modelo Poisson es mayor al del modelo Normal, lo que confirma lo antes expuesto.

Para los casos de tamaño de efecto mediano y grande se tiene la misma tendencia, en donde el modelo con mayor potencia es el Binomial negativo con un valor de 0.68 para el efecto mediano y 0.89 para el efecto grande. Se puede observar que la potencia aumenta

más en el cambio de efecto chico a mediano, que en el caso de mediano a grande, [Kyriazos y col. \(2018\)](#) comentan que el tamaño del efecto es una medida de la fuerza de la relación examinada. En la figura 4.20b se puede apreciar similitudes con el caso balanceado, donde las potencias tienden a aumentar conforme lo hace el tamaño del efecto; Se puede hacer hincapié en el hecho de que las potencias no son muy diferentes en el tamaño de efecto pequeño entre casos como se puede ver en la figura 4.20, para el tamaño de efecto mediano y grande son mayores en el caso balanceado; El modelo Normal tiende a ser mayor que el Normal transformado en el tamaño de efecto grande en el caso desbalanceado, a diferencia del caso balanceado donde la potencia es la misma.

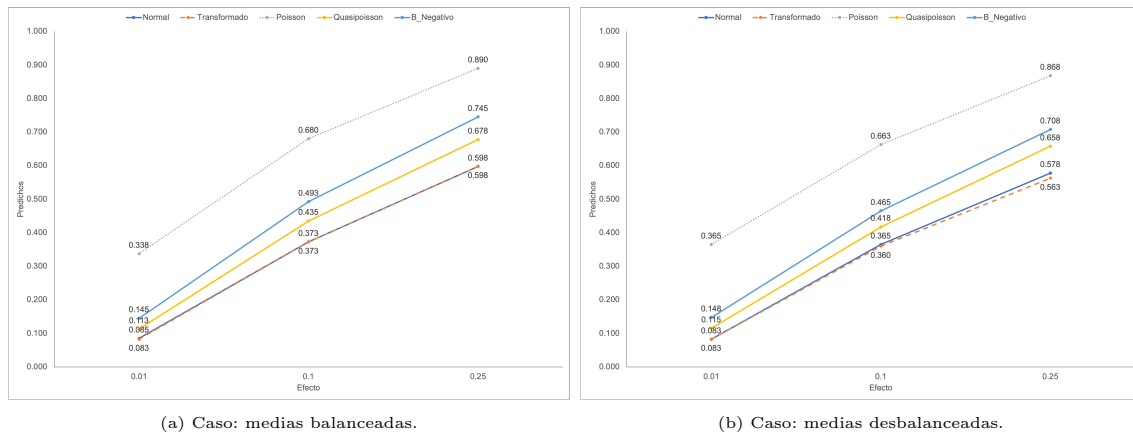


Figura 4.20. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través del tamaño del efecto.

Si analizamos la interacción para ocho tratamientos podemos ver en la figura 4.21 que, el comportamiento de los modelos es igual al de cuatro, podemos notar que la potencia aumenta conforme aumentan los tratamientos. Note que, el modelo Binomial negativo es el que presenta el incremento mayor.

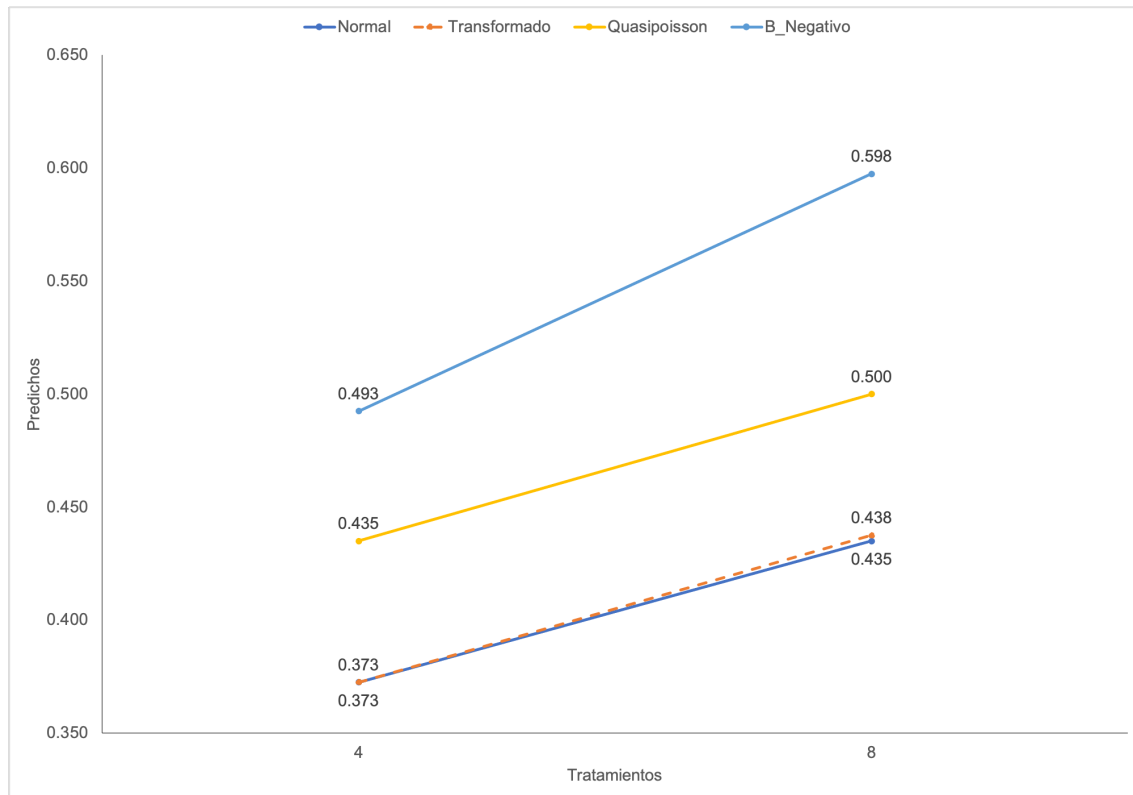


Figura 4.21. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de los tratamientos, caso de medias balanceadas y tamaño del efecto mediano.

Si analizamos la interacción entre las repeticiones, el tamaño del efecto y los modelos, como se ve representado en la figura 4.22 podemos ver que, entre menos repeticiones y menos tamaño de efecto, la potencia es menor, de manera similar, si las repeticiones y el tamaño del efecto son mayores, la potencia es mayor; Note que, como en el análisis anterior, los modelos Normal y Normal transformado son los que tienen la menor potencia en los diferentes tamaños de efecto y repeticiones, seguidos del Quasipoisson y el modelo que cuenta con la potencia mayor es el Binomial negativo; En el caso del efecto pequeño podemos observar en la figura 4.22a que, para las repeticiones seis y ocho se tiene la menor potencia seguido de cuatro y cincuenta repeticiones respectivamente; en los casos de tamaño de efecto mediano y grande, como se puede apreciar en las figuras 4.22b y 4.22c, a medida que las repeticiones aumentan, lo hace la potencia, teniendo un incremento mayor al pasar de ocho a cincuenta repeticiones.

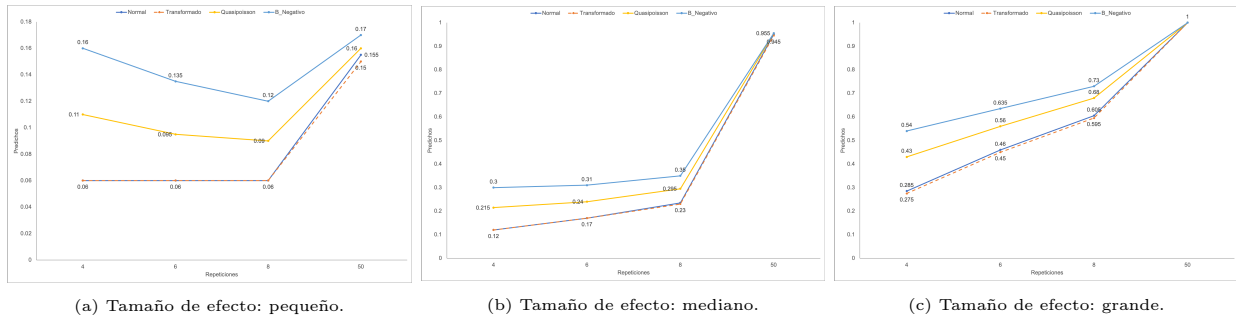


Figura 4.22. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de las repeticiones.

Esto mismo pasa cuando tenemos ocho tratamientos, pero la potencia aumenta al aumentar los tratamientos como podemos ver en la figura 4.23 para el caso exclusivo de ocho repeticiones.

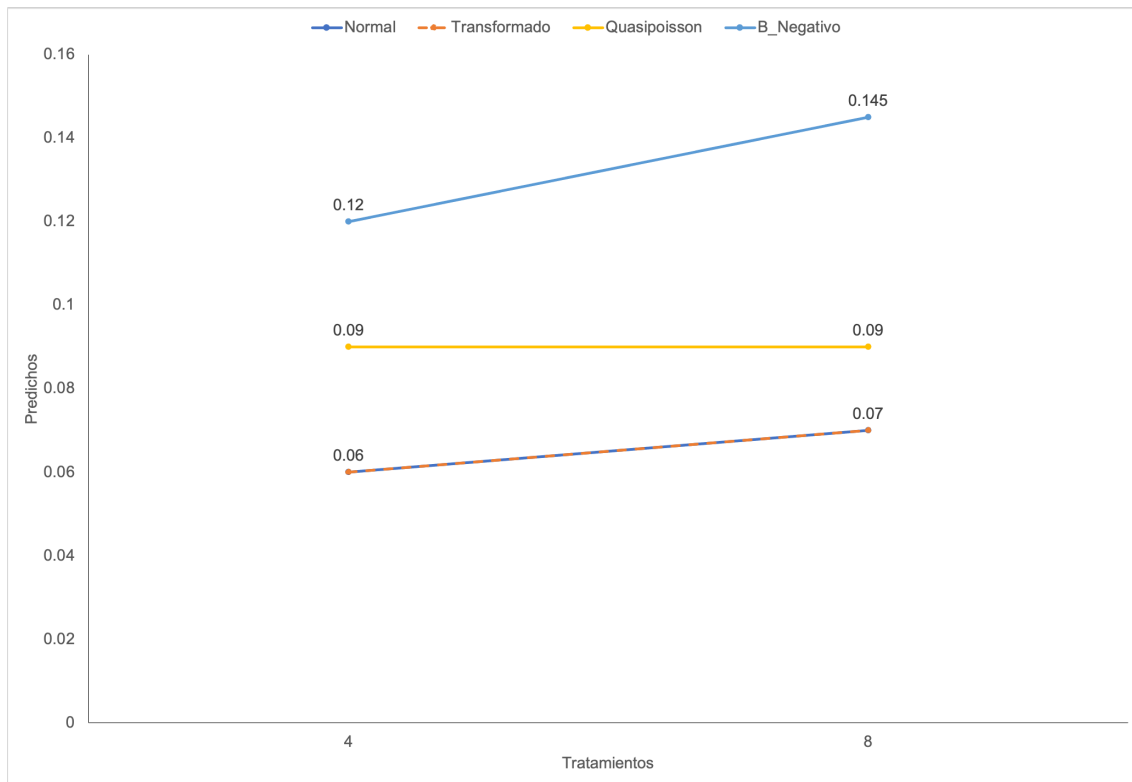


Figura 4.23. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión grande a través de los tratamientos, caso de ocho repeticiones y tamaño de efecto pequeño.

Al momento de analizar la interacción entre el caso, las repeticiones y el tamaño del efecto como lo muestra la figura 4.24, se aprecia que en el caso balanceado la potencia es mayor que en el caso desbalanceado, para cuando tenemos cuatro, seis y ocho repeticiones el cambio no es tan significativo al momento de pasar de un tamaño de efecto pequeño a uno grande, que cuando tenemos cincuenta repeticiones; En la figura 4.24a para cincuenta repeticiones, se

aprecia que no hay tanta diferencia entre los efectos mediano y grande; Para cuando tenemos el cambio de efecto pequeño a mediano, el aumento en la potencia es menor que el cambio entre el efecto mediano y el grande, esto mismo se aprecia en la figura 4.24b para el caso desbalanceado, con valores de la potencia menores.

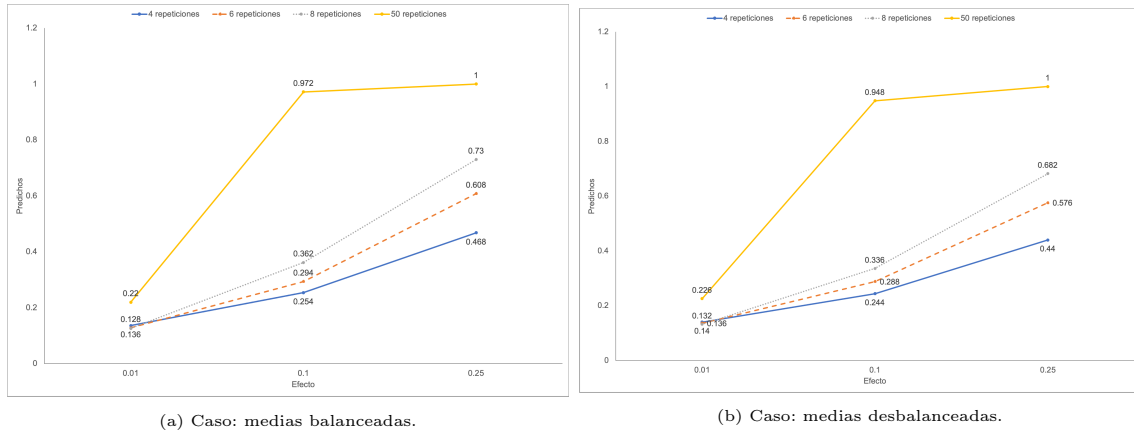


Figura 4.24. Comparación de la potencia de la prueba de las repeticiones para datos con sobredispersión grande a través de los tamaños del efecto.

Para ejemplificar la diferencia entre tratamientos, se utilizó el tamaño de efecto mediano, en donde las potencias aumentaban conforme aumentaban las repeticiones, note en la figura 4.25 que, las potencias son mayores cuando se utilizan ocho tratamientos, a cuando se utilizan cuatro.

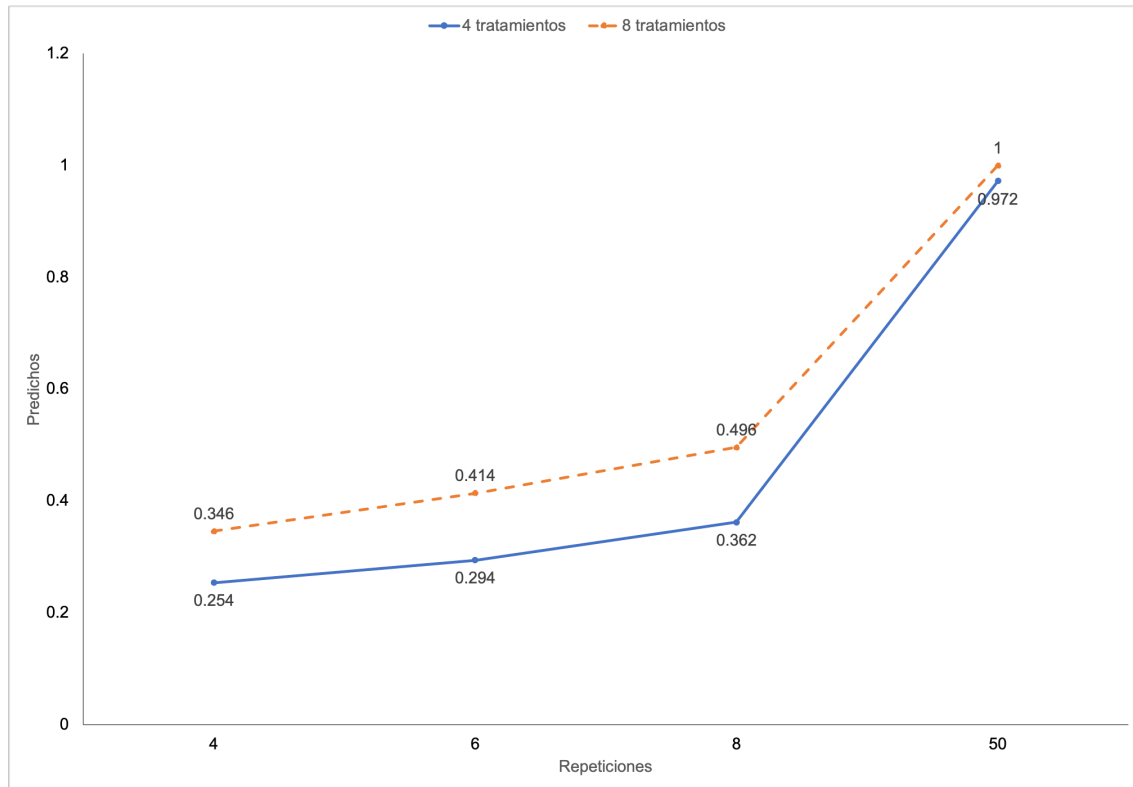
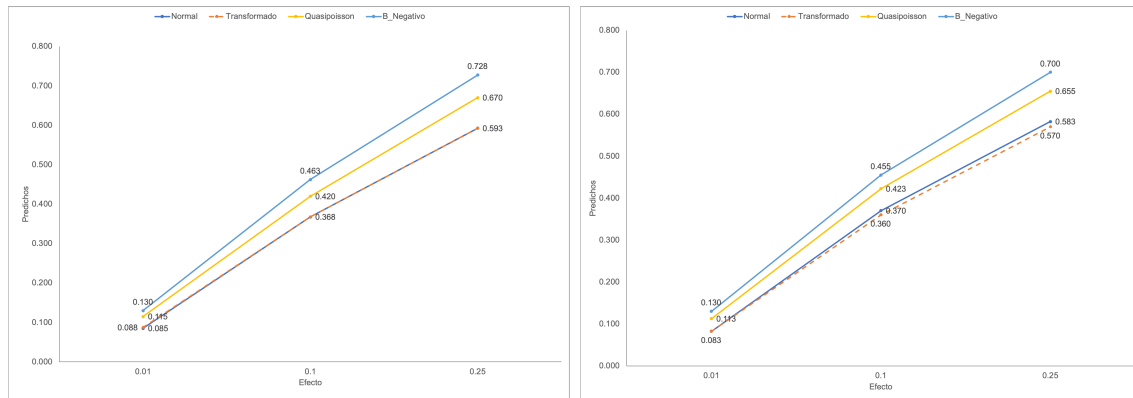


Figura 4.25. Comparación de la potencia de la prueba de los tratamientos para datos con sobredispersión grande a través de las repeticiones, caso de medias balanceadas y tamaño del efecto mediano.

4.2.3. Sobredispersión mediana

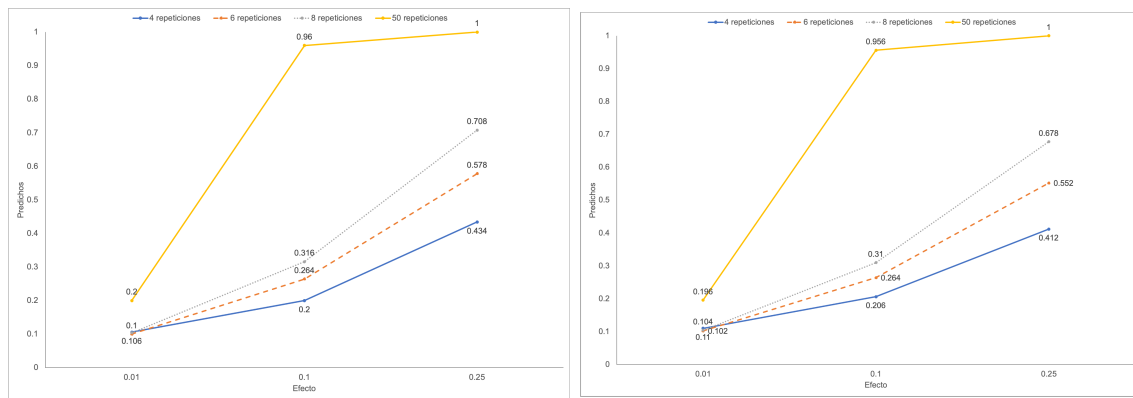
Para una sobredispersión mediana se separaron en dos, al igual que para datos sin sobredispersión y sobredispersión grande, cuando se utiliza cuatro y ocho tratamientos; el resultado en el modelo fueron dos interacciones triples: repetición, tamaño del efecto y modelo, junto con caso, tamaño del efecto y modelo, las cuales podemos ver en las figuras 4.26 y 4.27; En la figura 4.26a podemos apreciar que, al igual que trabajar con una sobredispersión grande, el modelo con la mayor potencia es el Binomial negativo seguido del Quasipoisson, y en igualdad de potencia, los modelos Normal y Normal transformado de manera general; Si vemos el tamaño de efecto pequeño se puede observar una ligera ganancia en potencia del modelo Normal transformado, la cuál se pierde en los otros tamaños de efecto. En la figura 4.26b se observa el mismo comportamiento de los modelos, pero la potencia tiende a ser ligeramente menor en la mayoría de los casos cuando utilizamos medias desbalanceadas; Note que para los modelos Normal y Normal transformado, no existen diferencias en los valores de las potencias.



(a) Caso: medias balanceadas.

(b) Caso: medias desbalanceadas.

Figura 4.26. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión mediana a través de los tamaño del efecto.



(a) Caso: medias balanceadas.

(b) Caso: medias desbalanceadas.

Figura 4.27. Comparación de la potencia de la prueba de las diferentes repeticiones para datos con sobredispersión mediana a través de los tamaño del efecto.

Para el análisis de la interacción caso, repetición y tamaño de efecto, si comparamos las figuras 4.27a y 4.27b se distingue que, los valores de las potencias son menores para el caso balanceado que para el caso desbalanceado. Como ya lo hemos resaltado en el caso de sobredispersión grande, el cambio entre el tamaño de efecto pequeño a mediano es mayor que el de mediano a grande, en ambos casos.

Cuando cambiamos de cuatro a ocho tratamientos, lo dicho anteriormente se mantiene, con la diferencia de que la potencia tiende a ser mayor cuando utilizamos ocho tratamientos como se advierte en la figura 4.28. Note que el cambio mayor lo tiene el modelo Binomial negativo, siendo este modelo, el que tiene una pendiente mayor en la gráfica.

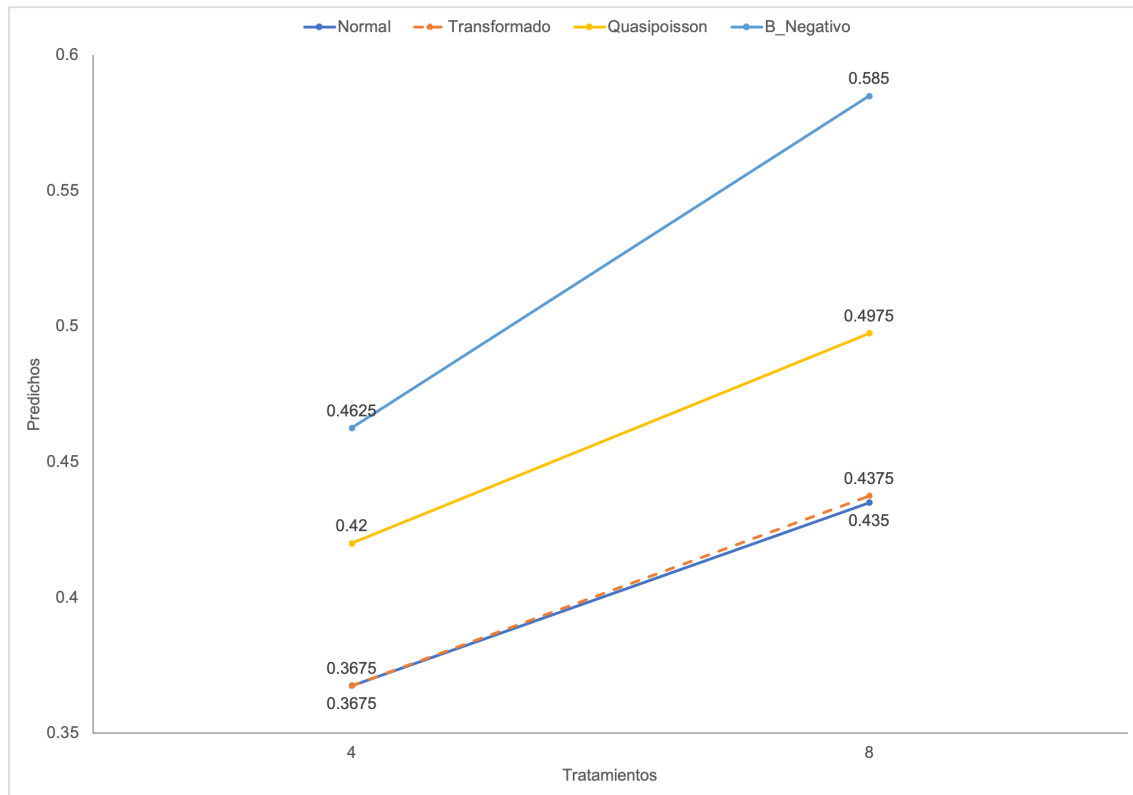


Figura 4.28. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión mediana a través de los tratamientos, caso de tamaño del efecto mediano y medias balanceadas.

Como se puede ver en la figura 4.29 cuando se utilizan cuatro, seis y ocho repeticiones, la potencia es muy diferente a cuando utilizamos cincuenta repeticiones, note también que, el cambio de las tres primeras repeticiones se mantiene alrededor de 0.1, a diferencia de cuando tenemos cincuenta, que el cambio ya es mínimo entre usar cuatro u ocho tratamientos.

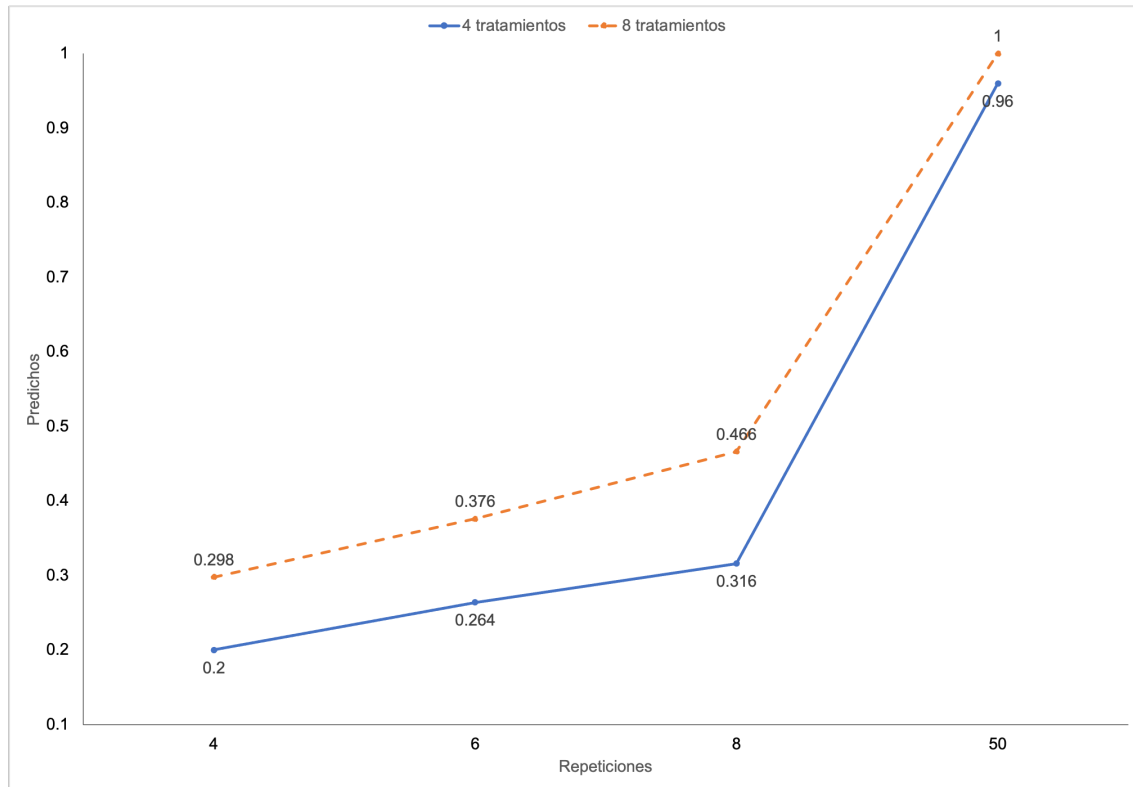


Figura 4.29. Comparación de la potencia de la prueba de los tratamientos para datos con sobredispersión mediana a través de las repeticiones, caso de tamaño del efecto mediano y medias balanceadas.

4.2.4. Sobredispersión pequeña

El análisis de la sobredispersión pequeña resultó de la misma manera que la sobredispersión mediana y grande, con ligeras diferencias, por lo que los análisis efectuados fueron de una manera general; se generaron dos interacciones triples: caso, repeticiones y tamaño del efecto y, repeticiones, tamaño del efecto y modelos. Para la segunda interacción, en la figura 4.30 se puede observar que a medida que el tamaño del efecto aumenta, lo hace la potencia; Los modelos Normal y Normal transformado tienen diferencias únicamente en el tamaño de efecto pequeño y, cuatro y cincuenta repeticiones, en los dos casos, el Normal transformado tiene una mayor potencia como lo muestra la figura 4.30a. Cuando el tamaño de efecto es grande y tenemos cincuenta repeticiones la potencia para todos los modelos es uno, como se aprecia en la figura 4.30c; Los incrementos en la potencia son mayores para cuatro, seis y ocho repeticiones cuando tenemos tamaño de efecto grande, teniendo para el caso de tamaño de efecto pequeño, que la potencia disminuye a medida que las repeticiones aumentan, en el cambio de ocho a cincuenta repeticiones, la potencia siempre aumenta. En todos los casos el modelo Binomial negativo es el que tiene mayor potencia, esto lo podemos ver en la figura 4.30b. Lo mismo sucede para ocho tratamientos, pero con una potencia mayor en todos los

casos.

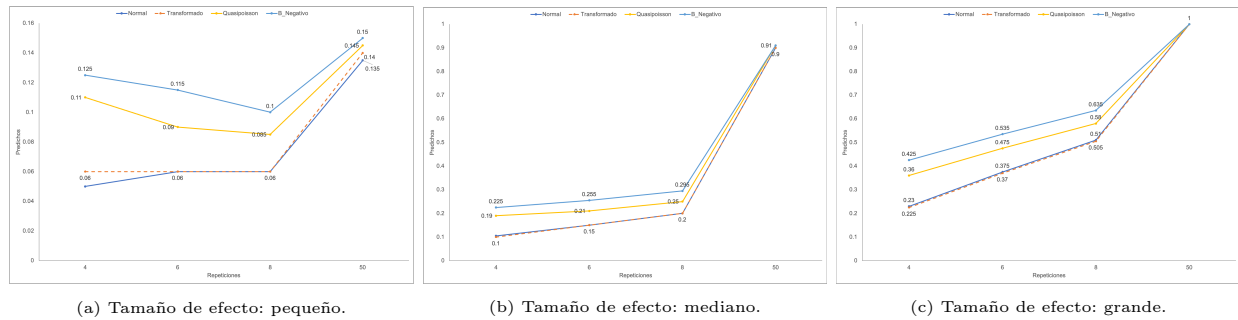


Figura 4.30. Comparación de la potencia de la prueba de los modelos para datos con sobredispersión pequeña a través de las repeticiones.

Para la interacción entre caso, repeticiones y tamaño del efecto tenemos en la figura 4.31 que, al igual que en las demás sobredispersiones, podemos ver que aumentando las repeticiones la potencia aumenta, en el caso del tamaño de efecto pequeño tenemos que la potencia casi es la misma para cuatro, seis y ocho repeticiones, siendo la única diferencia cuando tenemos cincuenta, esto pasa tanto para el caso balanceado como para el desbalanceado; El incremento entre tamaños de efecto se nota más en el caso balanceado como lo muestra la figura 4.31a, mientras que en el caso desbalanceado, el incremento en la potencia es menor como se aprecia en la figura 4.31b. Note que, el peso del tamaño del efecto es importante, al igual que la cantidad de repeticiones, es decir, si comparamos los valores de la potencia cuando tenemos cuatro repeticiones y un tamaño de efecto grande, a cuando lo comparamos con un tamaño de efecto pequeño y cincuenta repeticiones, podemos ver que el primer caso tiene un valor más grande; por otro lado, si comparamos dentro del mismo tamaño de efecto mediano, cuatro y cincuenta repeticiones, los valores de las potencias son muy diferentes, el valor de la potencia de cincuenta repeticiones es mayor de una manera significativa. En el caso de ocho tratamientos tenemos que las potencias aumentan de manera general.

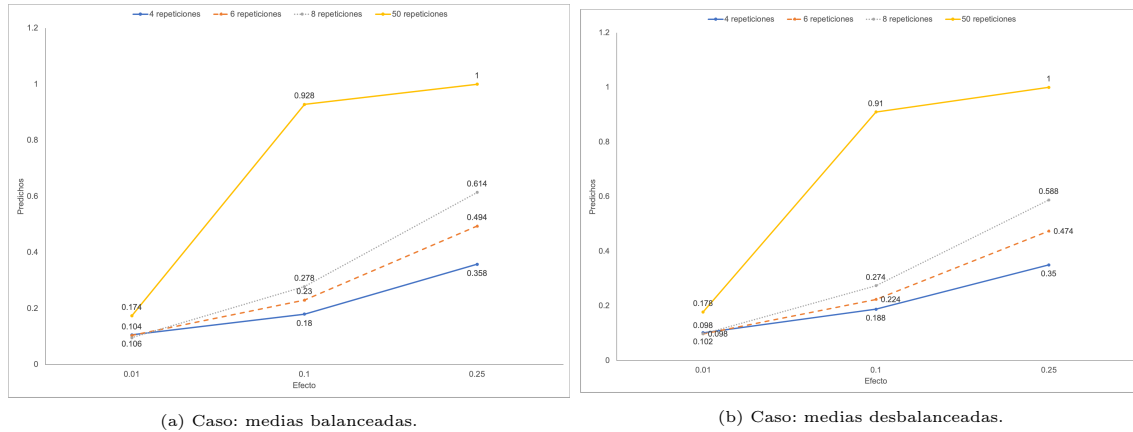


Figura 4.31. Comparación de la potencia de la prueba de las repeticiones para datos con sobredispersión pequeña a través del tamaño del efecto.

4.3. Prueba de sobredispersión

Podemos ver en la figura 4.32 que al momento de realizar la prueba de sobredispersión para el modelo GLM Poisson para el caso de las medias balanceadas, cuatro tratamientos y un tamaño de efecto pequeño, la prueba resulta deficiente, ya que, para los datos con sobredispersión se puede apreciar en 4.32a que, solamente los detecta por debajo de 0.18 cuando se tienen cuatro repeticiones, a diferencia de la figura 4.32b que es el caso de ocho repeticiones, los valores no son mayores a 0.53, lo cual, no es el caso para cincuenta repeticiones donde los valores de la prueba que se observan en la figura 4.32c son cercanos a uno cuando se tiene una sobredispersión pequeña o mediana, y uno cuando la sobredispersión es grande. Esto nos demuestra que no existe una forma confiable para saber si en los datos, existe o no sobredispersión, con tamaños de muestra pequeños.

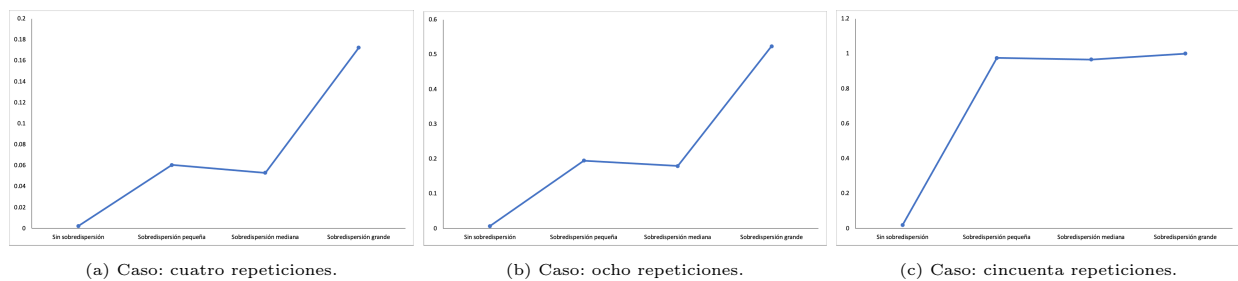


Figura 4.32. Comparación de la potencia de la prueba de sobredispersión para datos Poisson a través de las repeticiones, caso de medias balanceadas, cuatro tratamientos y tamaño de efecto pequeño.

Al momento de trabajar con cuatro u ocho tratamientos y haciendo la comparación, se puede ver en la figura 4.33 que, la prueba detecta más las sobredispersión cuando se tienen más tratamientos, note que, en los casos 4.33a y 4.33b, los valores no son mayores a 0.18 y

0.32 respectivamente, para el caso de medias balanceadas, cuatro repeticiones y un tamaño de efecto pequeño.

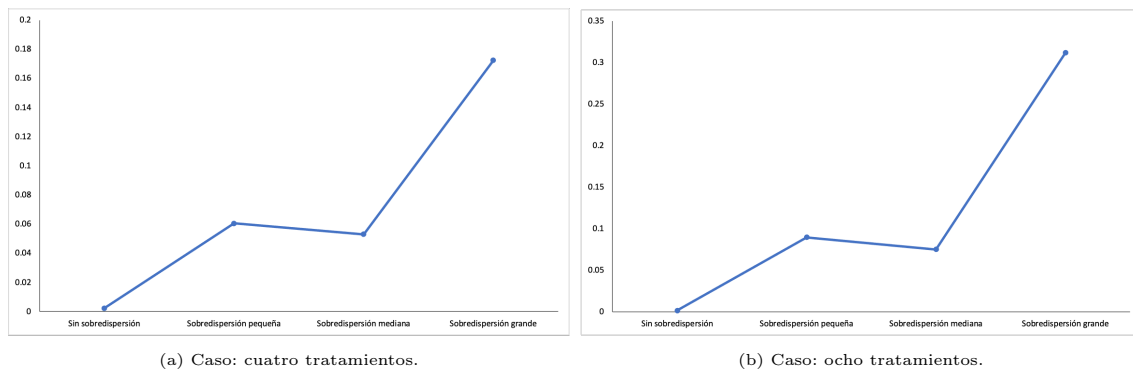


Figura 4.33. Comparación de la potencia de la prueba de sobredispersión para datos Poisson a través de los tratamientos, caso de medias balanceadas, cuatro repeticiones y tamaño de efecto pequeño.

5. CONCLUSIONES

Cuando los datos no presentan sobredispersión todos los enfoques mantuvieron el valor nominal de su tamaño de prueba, excepto el modelo Quasipoisson, el cual tuvo valores mayores a los nominales cuando se tenía una muestra pequeña, sólo para tamaño de muestra cincuenta mantuvo valores cercanos a los nominales. Con respecto a la potencia de la prueba, el modelo Poisson presentó la mayor.

En presencia de sobredispersión, el modelo Poisson no mantuvo el tamaño de prueba cercano al nominal, los modelos Binominal negativo y Quasipoisson sólo lo mantuvieron en muestras grandes (cincuenta repeticiones); en este caso los modelos Normal y Normal transformado mantuvieron el tamaño de la prueba cercano al nominal, sin embargo, en muestras grandes tuvieron menos potencia que el modelo Binominal negativo.

Cuando se tengan variables respuesta sin sobredispersión, se recomienda utilizar el modelo Poisson, en presencia de sobredispersión utilizar el modelo Normal o Normal transformado, cuando exista la posibilidad de tener cincuenta o más repeticiones, utilizar el modelo Binominal negativo.

Debido a los resultados obtenidos de la prueba de sobredispersión, y que el modelo Poisson y Normal de manera general tuvieron el mismo comportamiento, se recomienda utilizar el modelo Normal para todos los casos presentados.

6. LITERATURA CITADA

- Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.
- Bandera-Fernández, E. y Pérez-Pelea, L. (2018). Los modelos lineales generalizados mixtos. su aplicación en el mejoramiento de plantas. *Cultivos tropicales*, 39(1), 127–133.
- Box, G. E. y Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Cameron, A. C. y Trivedi, P. K. (1990). Regression-based tests for overdispersion in the poisson model. *Journal of econometrics*, 46(3), 347–364.
- Cameron, A. C. y Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.
- Casella, G. y Berger, R. L. (2001). *Statistical Inference*. Duxbury Press, 2 edition.
- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4), 417–451.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edn. á/l.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Cohen, J. (1994). The earth is round ($p < .05$). *American psychologist*, 49(12), 997.
- Condo, L. y Pazmiño, J. (2015). Diseño experimental en el desarrollo del conocimiento científico de las ciencias agropecuarias. *Tomo I*.
- Curtiss, J. H. (1943). On transformations used in the analysis of variance. *The Annals of Mathematical Statistics*, 14(2), 107–122.

- Declercq, L., Jamshidi, L., Fernández-Castilla, B., Beretvas, S. N., Moeyaert, M., Ferron, J. M., y Van den Noortgate, W. (2019). Analysis of single-case experimental count data using the linear mixed effects model: A simulation study. *Behavior Research Methods*, 51(6), 2477–2497.
- Faraway, J. J. (2015). *Linear models with R*. Texts in statistical science. CRC Press, second edition edition.
- Fisher, R. (1954). The analysis of variance with various binomial transformations. *Biometrics*, 10(1), 130–139.
- Fisher, R. A. (1970). Statistical methods for research workers. *Edinburgh: Oliver and Boyd*.
- Gabriel, J., Carlos, C., Valverde, A., y Indacochea, B. (2017). *Diseños experimentales: Teoría y práctica para experimentos agropecuarios*. Grupo Compás.
- Gardner, W., Mulvey, E. P., y Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological bulletin*, 118(3), 392.
- Gbur, E., of Agronomy, A. S., of America, S. S. S., y of America, C. S. S., Eds. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. Madison, WI: American Society of Agronomy: Soil Science Society of America: Crop Science Society of America. OCLC: ocn774487925.
- Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology*, 79(2), 314.
- Hilbe, J. M. (2017). The statistical analysis of count data/el análisis estadístico de los datos de recuento. *Cultura y Educación*, 29(3), 409–460.
- Hinkelmann, K. y Kempthorne, O. (1994). *Design and analysis of experiments*, volume 1. Wiley Online Library.
- Hrdličková, Z. (2006). Comparison of the power of the tests in one-way anova type model with poisson distributed variables. *Environmetrics*, 17, 227–237.
- Instituto SAS, G. d. u. (2011). *SAS/IML 9.3 user's guide*. Sas Institute.
- Ives, Anthony R.; Freckleton, R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution*, 6, 828–835.

- Jiménez, J. O. (2019). Comparación de los modelos lineal clásico y poisson desde la perspectiva de la aproximación poisson a la normal. *SERENQUETI*, (pp.9).
- Kempton, R. A., Fox, P. N., y Cerezo, M., Eds. (1996). *Statistical Methods for Plant Variety Evaluation*. Dordrecht: Springer Netherlands.
- Kyriazos, T. A. y others (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (efa, cfa) and sem in general. *Psychology*, 9(08), 2207.
- Lawson, J. (2014). *Design and Analysis of Experiments with R*, volume 115. CRC press.
- Lee, Y. y Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4), 619–656.
- Lewis, S. L., Montgomery, D. C., y Myers, R. H. (2001). Confidence interval coverage for designed experiments analyzed with glms. *Journal of quality technology*, 33(3), 279–292.
- Lipsey, M. W. (1990). *Design sensitivity*. Newbury Park, CA: Sage.
- Mahmud, M., Abrahamowicz, M., Leffondré, K., y Chaubey, Y. P. (2006). Selecting the optimal transformation of a continuous covariate in cox’s regression: Implications for hypothesis testing. *Communications in Statistics—Simulation and Computation*®, 35(1), 27–45.
- McArdle, B. H. y Anderson, M. J. (2004). Variance heterogeneity, transformations, and models of species abundance: a cautionary tale. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(7), 1294–1302.
- McCullagh, P. y Nelder, J. A. (1989). *Generalized linear models* 2nd edition chapman and hall. London, UK.
- Montgomery, D. C. (2004). *Diseño y análisis de experimentos*. 2º edición. limusa wiley.
- Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.
- Mood, A. M. (1950). *Introduction to the theory of statistics*.
- Mood, A. M., Graybill, F. A., Graybill, F. A., y Boes, D. C. (1974). *Introduction to the theory of statistics: McGraw-hill series in probability and statistics*.
- Murphy, K. R., Myers, B., y Wolach, A. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Routledge.

- Nelder, J. A. y Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370–384.
- Patnaik, P. (1949). The non-central χ^2 -and f-distribution and their applications. *Biometrika*, 36(1/2), 202–232.
- Qumsiyeh, M. y Shaughnessy, G. (2012). Comparison of re-sampling methods to generalized linear models and transformations in factorial and fractional factorial designs. *Journal of Modern Applied Statistical Methods*, 11(1), 95–105.
- Rosenthal, R. (1993). Cumulating evidence.
- Sedlmeier, P. y Gigerenzer, G. (1992). Do studies of statistical power have an effect on the power of studies?
- Self, S. G. y Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605–610.
- Shadish, W. R. y Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43, 971–980.
- Snedecor, G. W. y Cochran, W. G. (1989). Statistical methods, 8th edn. *Ames: Iowa State Univ. Press Iowa*, 54, 71–82.
- Steel, R. G. y others (1997). *Principles and procedures of statistics a biometrical approach*. Technical report.
- Tatsuoka, M. (1993). Elements of the general linear model.
- Thiemann, S. y Kraemer, H. (1987). How many subjects? statistical power analysis in research.
- Torres, V. y Seguí, Y. (2001). Procedimiento práctico para la determinación de la función de potencia a posteriori. *Revista Cubana de Ciencia Agrícola*, 35(4), 319–322.
- Wang, T., He, P., Ahn, K. W., Wang, X., Ghosh, S., y Laud, P. (2015). A re-formulation of generalized linear mixed models to fit family data in genetic association studies. *Frontiers in Genetics*, 6, 120.
- Warton, D. I., Lyons, M., Stoklosa, J., Ives, A. R., y Schielzeth, H. (2016). Three points to consider when choosing a lm or glm test for count data. *Methods in Ecology and Evolution*.

Young, L. J., Campbell, N. L., y Capuano, G. A. (1999). Analysis of overdispersed count data from single-factor experiments: A comparative study. *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 258–275.

ANEXOS

Tabla 5.1. Valores de las medias para datos Poisson.

Balanceadas			Desbalanceadas		
Tamaño del efecto					
Pequeño	Mediano	Grande	Pequeño	Mediano	Grande
Cuatro tratamientos					
18.0	20.2	20.4	23.7	21.7	27.4
18.0	20.2	20.4	22.7	18.6	21.6
18.7	22.9	25.4	22.7	18.6	21.6
18.7	22.9	25.4	22.7	18.6	21.6
Ocho tratamientos					
18.0	20.2	20.4	19.4	17.1	17.3
18.0	20.2	20.4	19.4	17.1	17.3
18.0	20.2	20.4	20.3	20.1	22.8
18.0	20.2	20.4	20.3	20.1	22.8
18.7	22.9	25.4	20.3	20.1	22.8
18.7	22.9	25.4	20.3	20.1	22.8
18.7	22.9	25.4	20.3	20.1	22.8
18.7	22.9	25.4	20.3	20.1	22.8

Tabla 5.2. Valores de las medias para datos con sobredispersión grande.

Balanceadas			Desbalanceadas		
Tamaño del efecto					
Pequeño	Mediano	Grande	Pequeño	Mediano	Grande
Cuatro tratamientos					
19.1	20.5	19.3	23.1	27.1	29.8
19.1	20.5	19.3	21.7	22.2	21.3
20.2	24.7	26.6	21.7	22.2	21.3
20.2	24.7	26.6	21.7	22.2	21.3
Ocho tratamientos					
18.3	20.5	19.2	19.2	17.8	15.0
18.3	20.5	19.2	19.2	17.8	15.0
18.3	20.5	19.2	20.6	22.4	23.0
18.3	20.5	19.2	20.6	22.4	23.0
19.4	24.8	26.8	20.6	22.4	23.0
19.4	24.8	26.8	20.6	22.4	23.0
19.4	24.8	26.8	20.6	22.4	23.0
19.4	24.8	26.8	20.6	22.4	23.0

Tabla 5.3. Valores de las medias para datos con sobredispersión mediana.

Balanceadas			Desbalanceadas		
Tamaño del efecto					
Pequeño	Mediano	Grande	Pequeño	Mediano	Grande
Cuatro tratamientos					
19.2	20.4	19.8	23.1	26.5	28.9
19.2	20.4	19.8	21.9	22.3	21.6
20.2	23.9	26.0	21.9	22.3	21.6
20.2	23.9	26.0	21.9	22.3	21.6
Ocho tratamientos					
18.4	20.7	19.7	19.1	18.3	15.9
18.4	20.7	19.7	19.1	18.3	15.9
18.4	20.7	19.7	20.3	22.4	22.8
18.4	20.7	19.7	20.3	22.4	22.8
19.3	24.4	26.2	20.3	22.4	22.8
19.3	24.4	26.2	20.3	22.4	22.8
19.3	24.4	26.2	20.3	22.4	22.8
19.3	24.4	26.2	20.3	22.4	22.8

Tabla 5.4. Valores de las medias para datos con sobredispersión pequeña.

Balanceadas		Desbalanceadas			
Tamaño del efecto					
Pequeño	Mediano	Grande	Pequeño	Mediano	Grande
Cuatro tratamientos					
20.3	19.6	18.7	20.2	24.2	24.9
20.3	19.6	18.7	19.2	20.6	18.9
21.2	22.7	24.0	19.2	20.6	18.9
21.2	22.7	24.0	19.2	20.6	18.9
Ocho tratamientos					
20.3	19.6	18.6	18.1	18.0	17.4
20.3	19.6	18.6	18.1	18.0	17.4
20.3	19.6	18.6	19.1	21.5	23.8
20.3	19.6	18.6	19.1	21.5	23.8
21.2	22.7	24.1	19.1	21.5	23.8
21.2	22.7	24.1	19.1	21.5	23.8
21.2	22.7	24.1	19.1	21.5	23.8
21.2	22.7	24.1	19.1	21.5	23.8

Tabla 5.5. Datos utilizados en la investigación.

Datos	Medias	Tratamiento	Repetición	Tamaño de efecto	Contador de sobredispersión
Sin sobredispersión	Balanceadas	4	4	0	0.0031
Sin sobredispersión	Balanceadas	4	4	0.01	0.0024
Sin sobredispersión	Balanceadas	4	4	0.1	0.0043
Sin sobredispersión	Balanceadas	4	4	0.25	0.0035
Sin sobredispersión	Balanceadas	4	6	0	0.0042
Sin sobredispersión	Balanceadas	4	6	0.01	0.0051
Sin sobredispersión	Balanceadas	4	6	0.1	0.0062
Sin sobredispersión	Balanceadas	4	6	0.25	0.0058
Sin sobredispersión	Balanceadas	4	8	0	0.0074
Sin sobredispersión	Balanceadas	4	8	0.01	0.0069
Sin sobredispersión	Balanceadas	4	8	0.1	0.0062
Sin sobredispersión	Balanceadas	4	8	0.25	0.0057
Sin sobredispersión	Balanceadas	4	50	0	0.0193
Sin sobredispersión	Balanceadas	4	50	0.01	0.0194
Sin sobredispersión	Balanceadas	4	50	0.1	0.0202
Sin sobredispersión	Balanceadas	4	50	0.25	0.0245
Sin sobredispersión	Desbalanceadas	4	4	0	0.0039
Sin sobredispersión	Desbalanceadas	4	4	0.01	0.0047
Sin sobredispersión	Desbalanceadas	4	4	0.1	0.0028
Sin sobredispersión	Desbalanceadas	4	4	0.25	0.0031
Sin sobredispersión	Desbalanceadas	4	6	0	0.0048
Sin sobredispersión	Desbalanceadas	4	6	0.01	0.0045
Sin sobredispersión	Desbalanceadas	4	6	0.1	0.0048
Sin sobredispersión	Desbalanceadas	4	6	0.25	0.005
Sin sobredispersión	Desbalanceadas	4	8	0	0.0044
Sin sobredispersión	Desbalanceadas	4	8	0.01	0.0068
Sin sobredispersión	Desbalanceadas	4	8	0.1	0.0053
Sin sobredispersión	Desbalanceadas	4	8	0.25	0.008
Sin sobredispersión	Desbalanceadas	4	50	0	0.0212

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.5. Datos utilizados en la investigación.

Datos	Medias	Tratamiento	Repetición	Tamaño de efecto	Contador de sobredispersión
Sin sobredispersión	Desbalanceadas	4	50	0.01	0.0195
Sin sobredispersión	Desbalanceadas	4	50	0.1	0.0216
Sin sobredispersión	Desbalanceadas	4	50	0.25	0.0224
Sin sobredispersión	Balanceadas	8	4	0	0.0017
Sin sobredispersión	Balanceadas	8	4	0.01	0.0015
Sin sobredispersión	Balanceadas	8	4	0.1	0.0016
Sin sobredispersión	Balanceadas	8	4	0.25	0.0009
Sin sobredispersión	Balanceadas	8	6	0	0.0038
Sin sobredispersión	Balanceadas	8	6	0.01	0.0022
Sin sobredispersión	Balanceadas	8	6	0.1	0.0029
Sin sobredispersión	Balanceadas	8	6	0.25	0.0044
Sin sobredispersión	Balanceadas	8	8	0	0.0049
Sin sobredispersión	Balanceadas	8	8	0.01	0.0044
Sin sobredispersión	Balanceadas	8	8	0.1	0.0043
Sin sobredispersión	Balanceadas	8	8	0.25	0.005
Sin sobredispersión	Balanceadas	8	50	0	0.0176
Sin sobredispersión	Balanceadas	8	50	0.01	0.016
Sin sobredispersión	Balanceadas	8	50	0.1	0.0216
Sin sobredispersión	Balanceadas	8	50	0.25	0.0218
Sin sobredispersión	Desbalanceadas	8	4	0	0.001
Sin sobredispersión	Desbalanceadas	8	4	0.01	0.0016
Sin sobredispersión	Desbalanceadas	8	4	0.1	0.0012
Sin sobredispersión	Desbalanceadas	8	4	0.25	0.0016
Sin sobredispersión	Desbalanceadas	8	6	0	0.0029
Sin sobredispersión	Desbalanceadas	8	6	0.01	0.0027
Sin sobredispersión	Desbalanceadas	8	6	0.1	0.003
Sin sobredispersión	Desbalanceadas	8	6	0.25	0.0032
Sin sobredispersión	Desbalanceadas	8	8	0	0.0043
Sin sobredispersión	Desbalanceadas	8	8	0.01	0.0041
Sin sobredispersión	Desbalanceadas	8	8	0.1	0.0047
Sin sobredispersión	Desbalanceadas	8	8	0.25	0.0041
Sin sobredispersión	Desbalanceadas	8	50	0	0.0193
Sin sobredispersión	Desbalanceadas	8	50	0.01	0.0205
Sin sobredispersión	Desbalanceadas	8	50	0.1	0.0165
Sin sobredispersión	Desbalanceadas	8	50	0.25	0.019
Sobredispersión grande	Balanceadas	4	4	0	0.1869
Sobredispersión grande	Balanceadas	4	4	0.01	0.0531
Sobredispersión grande	Balanceadas	4	4	0.1	0.2203
Sobredispersión grande	Balanceadas	4	4	0.25	0.2244
Sobredispersión grande	Balanceadas	4	6	0	0.354
Sobredispersión grande	Balanceadas	4	6	0.01	0.3487
Sobredispersión grande	Balanceadas	4	6	0.1	0.4393
Sobredispersión grande	Balanceadas	4	6	0.25	0.434
Sobredispersión grande	Balanceadas	4	8	0	0.5344
Sobredispersión grande	Balanceadas	4	8	0.01	0.195
Sobredispersión grande	Balanceadas	4	8	0.1	0.6174
Sobredispersión grande	Balanceadas	4	8	0.25	0.6279
Sobredispersión grande	Balanceadas	4	50	0	1
Sobredispersión grande	Balanceadas	4	50	0.01	1
Sobredispersión grande	Balanceadas	4	50	0.1	1
Sobredispersión grande	Balanceadas	4	50	0.25	1
Sobredispersión grande	Desbalanceadas	4	4	0	0.1853
Sobredispersión grande	Desbalanceadas	4	4	0.01	0.2189
Sobredispersión grande	Desbalanceadas	4	4	0.1	0.243
Sobredispersión grande	Desbalanceadas	4	4	0.25	0.2475
Sobredispersión grande	Desbalanceadas	4	6	0	0.3619
Sobredispersión grande	Desbalanceadas	4	6	0.01	0.4259
Sobredispersión grande	Desbalanceadas	4	6	0.1	0.4531
Sobredispersión grande	Desbalanceadas	4	6	0.25	0.4449
Sobredispersión grande	Desbalanceadas	4	8	0	0.5331
Sobredispersión grande	Desbalanceadas	4	8	0.01	0.6101
Sobredispersión grande	Desbalanceadas	4	8	0.1	0.6444

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.5. Datos utilizados en la investigación.

Datos	Medias	Tratamiento	Repetición	Tamaño de efecto	Contador de sobredispersión
Sobredispersión grande	Desbalanceadas	4	8	0.25	0.6349
Sobredispersión grande	Desbalanceadas	4	50	0	1
Sobredispersión grande	Desbalanceadas	4	50	0.01	1
Sobredispersión grande	Desbalanceadas	4	50	0.1	1
Sobredispersión grande	Desbalanceadas	4	50	0.25	1
Sobredispersión grande	Balanceadas	8	4	0	0.3385
Sobredispersión grande	Balanceadas	8	4	0.01	0.3118
Sobredispersión grande	Balanceadas	8	4	0.1	0.4232
Sobredispersión grande	Balanceadas	8	4	0.25	0.4342
Sobredispersión grande	Balanceadas	8	6	0	0.6504
Sobredispersión grande	Balanceadas	8	6	0.01	0.6081
Sobredispersión grande	Balanceadas	8	6	0.1	0.7378
Sobredispersión grande	Balanceadas	8	6	0.25	0.7545
Sobredispersión grande	Balanceadas	8	8	0	0.848
Sobredispersión grande	Balanceadas	8	8	0.01	0.3695
Sobredispersión grande	Balanceadas	8	8	0.1	0.8988
Sobredispersión grande	Balanceadas	8	8	0.25	0.9071
Sobredispersión grande	Balanceadas	8	50	0	1
Sobredispersión grande	Balanceadas	8	50	0.01	1
Sobredispersión grande	Balanceadas	8	50	0.1	1
Sobredispersión grande	Balanceadas	8	50	0.25	1
Sobredispersión grande	Desbalanceadas	8	4	0	0.3403
Sobredispersión grande	Desbalanceadas	8	4	0.01	0.3472
Sobredispersión grande	Desbalanceadas	8	4	0.1	0.3758
Sobredispersión grande	Desbalanceadas	8	4	0.25	0.3734
Sobredispersión grande	Desbalanceadas	8	6	0	0.6482
Sobredispersión grande	Desbalanceadas	8	6	0.01	0.6641
Sobredispersión grande	Desbalanceadas	8	6	0.1	0.6985
Sobredispersión grande	Desbalanceadas	8	6	0.25	0.6888
Sobredispersión grande	Desbalanceadas	8	8	0	0.8414
Sobredispersión grande	Desbalanceadas	8	8	0.01	0.8544
Sobredispersión grande	Desbalanceadas	8	8	0.1	0.8803
Sobredispersión grande	Desbalanceadas	8	8	0.25	0.873
Sobredispersión grande	Desbalanceadas	8	50	0	1
Sobredispersión grande	Desbalanceadas	8	50	0.01	1
Sobredispersión grande	Desbalanceadas	8	50	0.1	1
Sobredispersión grande	Desbalanceadas	8	50	0.25	1
Sobredispersión mediana	Balanceadas	4	4	0	0.0573
Sobredispersión mediana	Balanceadas	4	4	0.01	0.0606
Sobredispersión mediana	Balanceadas	4	4	0.1	0.0619
Sobredispersión mediana	Balanceadas	4	4	0.25	0.0702
Sobredispersión mediana	Balanceadas	4	6	0	0.1129
Sobredispersión mediana	Balanceadas	4	6	0.01	0.1122
Sobredispersión mediana	Balanceadas	4	6	0.1	0.1352
Sobredispersión mediana	Balanceadas	4	6	0.25	0.1447
Sobredispersión mediana	Balanceadas	4	8	0	0.1827
Sobredispersión mediana	Balanceadas	4	8	0.01	0.1794
Sobredispersión mediana	Balanceadas	4	8	0.1	0.2149
Sobredispersión mediana	Balanceadas	4	8	0.25	0.229
Sobredispersión mediana	Balanceadas	4	50	0	0.9707
Sobredispersión mediana	Balanceadas	4	50	0.01	0.9663
Sobredispersión mediana	Balanceadas	4	50	0.1	0.9862
Sobredispersión mediana	Balanceadas	4	50	0.25	0.9899
Sobredispersión mediana	Desbalanceadas	4	4	0	0.0542
Sobredispersión mediana	Desbalanceadas	4	4	0.01	0.0659
Sobredispersión mediana	Desbalanceadas	4	4	0.1	0.0696
Sobredispersión mediana	Desbalanceadas	4	4	0.25	0.0728
Sobredispersión mediana	Desbalanceadas	4	6	0	0.1111
Sobredispersión mediana	Desbalanceadas	4	6	0.01	0.1344
Sobredispersión mediana	Desbalanceadas	4	6	0.1	0.1421
Sobredispersión mediana	Desbalanceadas	4	6	0.25	0.1541
Sobredispersión mediana	Desbalanceadas	4	8	0	0.1886

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.5. Datos utilizados en la investigación.

Datos	Medias	Tratamiento	Repetición	Tamaño de efecto	Contador de sobredispersión
Sobredispersión mediana	Desbalanceadas	4	8	0.01	0.2283
Sobredispersión mediana	Desbalanceadas	4	8	0.1	0.2432
Sobredispersión mediana	Desbalanceadas	4	8	0.25	0.2357
Sobredispersión mediana	Desbalanceadas	4	50	0	0.9704
Sobredispersión mediana	Desbalanceadas	4	50	0.01	0.9876
Sobredispersión mediana	Desbalanceadas	4	50	0.1	0.9921
Sobredispersión mediana	Desbalanceadas	4	50	0.25	0.993
Sobredispersión mediana	Balanceadas	8	4	0	0.0848
Sobredispersión mediana	Balanceadas	8	4	0.01	0.075
Sobredispersión mediana	Balanceadas	8	4	0.1	0.1053
Sobredispersión mediana	Balanceadas	8	4	0.25	0.1131
Sobredispersión mediana	Balanceadas	8	6	0	0.2074
Sobredispersión mediana	Balanceadas	8	6	0.01	0.1844
Sobredispersión mediana	Balanceadas	8	6	0.1	0.2573
Sobredispersión mediana	Balanceadas	8	6	0.25	0.2619
Sobredispersión mediana	Balanceadas	8	8	0	0.3549
Sobredispersión mediana	Balanceadas	8	8	0.01	0.307
Sobredispersión mediana	Balanceadas	8	8	0.1	0.4212
Sobredispersión mediana	Balanceadas	8	8	0.25	0.4466
Sobredispersión mediana	Balanceadas	8	50	0	0.9995
Sobredispersión mediana	Balanceadas	8	50	0.01	0.9992
Sobredispersión mediana	Balanceadas	8	50	0.1	0.9999
Sobredispersión mediana	Balanceadas	8	50	0.25	1
Sobredispersión mediana	Desbalanceadas	8	4	0	0.0774
Sobredispersión mediana	Desbalanceadas	8	4	0.01	0.0831
Sobredispersión mediana	Desbalanceadas	8	4	0.1	0.0929
Sobredispersión mediana	Desbalanceadas	8	4	0.25	0.0895
Sobredispersión mediana	Desbalanceadas	8	6	0	0.2022
Sobredispersión mediana	Desbalanceadas	8	6	0.01	0.2065
Sobredispersión mediana	Desbalanceadas	8	6	0.1	0.2379
Sobredispersión mediana	Desbalanceadas	8	6	0.25	0.227
Sobredispersión mediana	Desbalanceadas	8	8	0	0.3523
Sobredispersión mediana	Desbalanceadas	8	8	0.01	0.3488
Sobredispersión mediana	Desbalanceadas	8	8	0.1	0.3908
Sobredispersión mediana	Desbalanceadas	8	8	0.25	0.385
Sobredispersión mediana	Desbalanceadas	8	50	0	0.9997
Sobredispersión mediana	Desbalanceadas	8	50	0.01	0.9995
Sobredispersión mediana	Desbalanceadas	8	50	0.1	0.9999
Sobredispersión mediana	Desbalanceadas	8	50	0.25	0.9999
Sobredispersión pequeña	Balanceadas	4	4	0	0.0551
Sobredispersión pequeña	Balanceadas	4	4	0.01	0.1724
Sobredispersión pequeña	Balanceadas	4	4	0.1	0.0626
Sobredispersión pequeña	Balanceadas	4	4	0.25	0.063
Sobredispersión pequeña	Balanceadas	4	6	0	0.1114
Sobredispersión pequeña	Balanceadas	4	6	0.01	0.1165
Sobredispersión pequeña	Balanceadas	4	6	0.1	0.1194
Sobredispersión pequeña	Balanceadas	4	6	0.25	0.1207
Sobredispersión pequeña	Balanceadas	4	8	0	0.1866
Sobredispersión pequeña	Balanceadas	4	8	0.01	0.5236
Sobredispersión pequeña	Balanceadas	4	8	0.1	0.1992
Sobredispersión pequeña	Balanceadas	4	8	0.25	0.2051
Sobredispersión pequeña	Balanceadas	4	50	0	0.9667
Sobredispersión pequeña	Balanceadas	4	50	0.01	0.9759
Sobredispersión pequeña	Balanceadas	4	50	0.1	0.9818
Sobredispersión pequeña	Balanceadas	4	50	0.25	0.9807
Sobredispersión pequeña	Desbalanceadas	4	4	0	0.0574
Sobredispersión pequeña	Desbalanceadas	4	4	0.01	0.0526
Sobredispersión pequeña	Desbalanceadas	4	4	0.1	0.0665
Sobredispersión pequeña	Desbalanceadas	4	4	0.25	0.0569
Sobredispersión pequeña	Desbalanceadas	4	6	0	0.1099
Sobredispersión pequeña	Desbalanceadas	4	6	0.01	0.1053
Sobredispersión pequeña	Desbalanceadas	4	6	0.1	0.1285

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.5. Datos utilizados en la investigación.

Datos	Medias	Tratamiento	Repetición	Tamaño de efecto	Contador de sobredispersión
Sobredispersión pequeña	Desbalanceadas	4	6	0.25	0.1165
Sobredispersión pequeña	Desbalanceadas	4	8	0	0.1804
Sobredispersión pequeña	Desbalanceadas	4	8	0.01	0.1711
Sobredispersión pequeña	Desbalanceadas	4	8	0.1	0.2128
Sobredispersión pequeña	Desbalanceadas	4	8	0.25	0.1943
Sobredispersión pequeña	Desbalanceadas	4	50	0	0.9687
Sobredispersión pequeña	Desbalanceadas	4	50	0.01	0.9654
Sobredispersión pequeña	Desbalanceadas	4	50	0.1	0.9827
Sobredispersión pequeña	Desbalanceadas	4	50	0.25	0.9752
Sobredispersión pequeña	Balanceadas	8	4	0	0.0815
Sobredispersión pequeña	Balanceadas	8	4	0.01	0.0894
Sobredispersión pequeña	Balanceadas	8	4	0.1	0.0931
Sobredispersión pequeña	Balanceadas	8	4	0.25	0.0909
Sobredispersión pequeña	Balanceadas	8	6	0	0.1995
Sobredispersión pequeña	Balanceadas	8	6	0.01	0.2198
Sobredispersión pequeña	Balanceadas	8	6	0.1	0.2351
Sobredispersión pequeña	Balanceadas	8	6	0.25	0.2377
Sobredispersión pequeña	Balanceadas	8	8	0	0.3498
Sobredispersión pequeña	Balanceadas	8	8	0.01	0.8112
Sobredispersión pequeña	Balanceadas	8	8	0.1	0.3816
Sobredispersión pequeña	Balanceadas	8	8	0.25	0.3922
Sobredispersión pequeña	Balanceadas	8	50	0	0.9998
Sobredispersión pequeña	Balanceadas	8	50	0.01	1
Sobredispersión pequeña	Balanceadas	8	50	0.1	1
Sobredispersión pequeña	Balanceadas	8	50	0.25	0.9998
Sobredispersión pequeña	Desbalanceadas	8	4	0	0.0845
Sobredispersión pequeña	Desbalanceadas	8	4	0.01	0.0727
Sobredispersión pequeña	Desbalanceadas	8	4	0.1	0.0905
Sobredispersión pequeña	Desbalanceadas	8	4	0.25	0.1015
Sobredispersión pequeña	Desbalanceadas	8	6	0	0.209
Sobredispersión pequeña	Desbalanceadas	8	6	0.01	0.1763
Sobredispersión pequeña	Desbalanceadas	8	6	0.1	0.2243
Sobredispersión pequeña	Desbalanceadas	8	6	0.25	0.2555
Sobredispersión pequeña	Desbalanceadas	8	8	0	0.3428
Sobredispersión pequeña	Desbalanceadas	8	8	0.01	0.3054
Sobredispersión pequeña	Desbalanceadas	8	8	0.1	0.3545
Sobredispersión pequeña	Desbalanceadas	8	8	0.25	0.4094
Sobredispersión pequeña	Desbalanceadas	8	50	0	0.9997
Sobredispersión pequeña	Desbalanceadas	8	50	0.01	0.9997
Sobredispersión pequeña	Desbalanceadas	8	50	0.1	0.9998
Sobredispersión pequeña	Desbalanceadas	8	50	0.25	1

(Fin de la tabla.)

Tabla 5.6. Valores de potencia de los modelos utilizados en la investigación.

Datos	Normal	Normal transformado	Poisson	Quasipoisson	Binomial negativo
Sin sobredispersión	0.0555	0.0553	0.0556	0.1041	0.0498
Sin sobredispersión	0.053	0.0527	0.0553	0.1074	0.0508
Sin sobredispersión	0.1209	0.1219	0.1501	0.2058	0.1401
Sin sobredispersión	0.2866	0.2872	0.3931	0.4325	0.373
Sin sobredispersión	0.0467	0.0464	0.0484	0.0774	0.0425
Sin sobredispersión	0.0575	0.0567	0.0596	0.0899	0.0534
Sin sobredispersión	0.1672	0.1673	0.191	0.2299	0.1783
Sin sobredispersión	0.4627	0.4635	0.5504	0.5629	0.5255
Sin sobredispersión	0.0506	0.05	0.05	0.0705	0.0453
Sin sobredispersión	0.0591	0.0598	0.0633	0.0852	0.0557
Sin sobredispersión	0.2216	0.2193	0.2473	0.2766	0.2293
Sin sobredispersión	0.6253	0.6227	0.695	0.6895	0.6693
Sin sobredispersión	0.0492	0.0506	0.0493	0.0525	0.0451

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.6. Valores de potencia de los modelos utilizados en la investigación.

Datos	Normal	Normal transformado	Poisson	Quasipoisson	Binomial negativo
Sin sobredispersión	0.1394	0.1384	0.1418	0.1435	0.1311
Sin sobredispersión	0.9454	0.9443	0.9501	0.9479	0.9437
Sin sobredispersión	1	1	1	1	1
Sin sobredispersión	0.0529	0.0527	0.0508	0.103	0.0471
Sin sobredispersión	0.0557	0.0553	0.0582	0.1101	0.0534
Sin sobredispersión	0.1205	0.1162	0.1469	0.2096	0.1351
Sin sobredispersión	0.2904	0.2784	0.3818	0.4286	0.3626
Sin sobredispersión	0.0474	0.0469	0.0465	0.0766	0.041
Sin sobredispersión	0.0593	0.0585	0.0603	0.0916	0.0528
Sin sobredispersión	0.1854	0.1815	0.2174	0.256	0.1989
Sin sobredispersión	0.4669	0.4487	0.5402	0.5627	0.5161
Sin sobredispersión	0.049	0.0495	0.0479	0.0684	0.0424
Sin sobredispersión	0.0642	0.0642	0.0674	0.0898	0.0597
Sin sobredispersión	0.2398	0.2306	0.2756	0.296	0.2558
Sin sobredispersión	0.6195	0.602	0.6784	0.6801	0.6513
Sin sobredispersión	0.0494	0.0488	0.049	0.0531	0.0446
Sin sobredispersión	0.1668	0.1628	0.1638	0.1731	0.1554
Sin sobredispersión	0.9559	0.9521	0.9582	0.9583	0.9537
Sin sobredispersión	1	1	1	1	1
Sin sobredispersión	0.0486	0.0489	0.0482	0.093	0.0453
Sin sobredispersión	0.054	0.0546	0.0568	0.1048	0.0539
Sin sobredispersión	0.1433	0.1438	0.1793	0.2358	0.1733
Sin sobredispersión	0.4138	0.4136	0.5435	0.5502	0.5292
Sin sobredispersión	0.0526	0.0525	0.0536	0.0808	0.0495
Sin sobredispersión	0.0571	0.0571	0.0611	0.0882	0.0571
Sin sobredispersión	0.2156	0.2137	0.2593	0.2861	0.2463
Sin sobredispersión	0.6654	0.6658	0.7543	0.7418	0.7383
Sin sobredispersión	0.0501	0.0497	0.0506	0.0686	0.0461
Sin sobredispersión	0.067	0.067	0.0695	0.0944	0.0639
Sin sobredispersión	0.3054	0.3012	0.3373	0.3615	0.3222
Sin sobredispersión	0.8326	0.8296	0.8861	0.8721	0.8756
Sin sobredispersión	0.0479	0.0473	0.0466	0.05	0.043
Sin sobredispersión	0.1784	0.1779	0.1814	0.1853	0.171
Sin sobredispersión	0.9954	0.9951	0.9957	0.9957	0.9952
Sin sobredispersión	1	1	1	1	1
Sin sobredispersión	0.0464	0.0468	0.049	0.0909	0.0459
Sin sobredispersión	0.0558	0.0557	0.059	0.1052	0.0554
Sin sobredispersión	0.139	0.1437	0.179	0.23	0.1718
Sin sobredispersión	0.4079	0.4303	0.5595	0.5474	0.5465
Sin sobredispersión	0.0467	0.0475	0.0492	0.0747	0.0461
Sin sobredispersión	0.061	0.0622	0.0626	0.0944	0.0584
Sin sobredispersión	0.2214	0.2287	0.2631	0.2855	0.2491
Sin sobredispersión	0.6594	0.6853	0.7693	0.7441	0.7537
Sin sobredispersión	0.0495	0.0487	0.0504	0.0687	0.0456
Sin sobredispersión	0.0689	0.0696	0.0713	0.0912	0.0656
Sin sobredispersión	0.3111	0.3211	0.3626	0.3693	0.3446
Sin sobredispersión	0.8337	0.8494	0.8945	0.8713	0.8838
Sin sobredispersión	0.0509	0.0511	0.0504	0.0543	0.0468
Sin sobredispersión	0.1926	0.1967	0.196	0.1998	0.1869
Sin sobredispersión	0.9973	0.9977	0.9979	0.9976	0.9977
Sin sobredispersión	1	1	1	1	1
Sobredispersión grande	0.0497	0.0508	0.2751	0.1042	0.1526
Sobredispersión grande	0.0608	0.0597	0.1775	0.1149	0.129
Sobredispersión grande	0.1224	0.1229	0.5028	0.2171	0.3123
Sobredispersión grande	0.2903	0.2958	0.7604	0.4348	0.5601
Sobredispersión grande	0.0508	0.0507	0.2727	0.0798	0.1169
Sobredispersión grande	0.0605	0.0614	0.2925	0.0965	0.1359
Sobredispersión grande	0.1663	0.1694	0.5728	0.239	0.3164
Sobredispersión grande	0.4686	0.474	0.87	0.5741	0.663
Sobredispersión grande	0.0511	0.0523	0.2757	0.0734	0.1024
Sobredispersión grande	0.0578	0.059	0.1836	0.0816	0.1054
Sobredispersión grande	0.2422	0.2424	0.6474	0.3075	0.3675

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.6. Valores de potencia de los modelos utilizados en la investigación.

Datos	Normal	Normal transformado	Poisson	Quasipoisson	Binomial negativo
Sobredispersión grande	0.6291	0.6313	0.9334	0.7031	0.7628
Sobredispersión grande	0.0495	0.0499	0.273	0.0533	0.0566
Sobredispersión grande	0.1521	0.1524	0.4724	0.1604	0.1679
Sobredispersión grande	0.9619	0.9595	0.9978	0.965	0.9678
Sobredispersión grande	1	1	1	1	1
Sobredispersión grande	0.0502	0.0517	0.2766	0.1035	0.1517
Sobredispersión grande	0.0548	0.0534	0.3083	0.1118	0.1615
Sobredispersión grande	0.1181	0.116	0.4831	0.2057	0.2873
Sobredispersión grande	0.278	0.2614	0.7264	0.4208	0.5245
Sobredispersión grande	0.0488	0.0491	0.2794	0.081	0.1172
Sobredispersión grande	0.0564	0.0579	0.3281	0.094	0.1353
Sobredispersión grande	0.1745	0.1688	0.5597	0.2362	0.3029
Sobredispersión grande	0.4494	0.4276	0.8337	0.5479	0.6128
Sobredispersión grande	0.0511	0.0507	0.2612	0.0743	0.1012
Sobredispersión grande	0.0602	0.0604	0.3291	0.0871	0.1164
Sobredispersión grande	0.2266	0.2186	0.622	0.2841	0.3363
Sobredispersión grande	0.5859	0.5668	0.9038	0.6546	0.7
Sobredispersión grande	0.0491	0.0497	0.2727	0.0532	0.0558
Sobredispersión grande	0.1582	0.1517	0.4873	0.1655	0.1704
Sobredispersión grande	0.9374	0.9322	0.994	0.9407	0.9402
Sobredispersión grande	1	1	1	1	1
Sobredispersión grande	0.0488	0.0489	0.4324	0.0985	0.204
Sobredispersión grande	0.06	0.0599	0.4255	0.1158	0.217
Sobredispersión grande	0.1605	0.1616	0.7238	0.2634	0.4318
Sobredispersión grande	0.4401	0.4449	0.9424	0.5963	0.766
Sobredispersión grande	0.0493	0.0506	0.4188	0.0777	0.1409
Sobredispersión grande	0.067	0.0674	0.4481	0.0994	0.1711
Sobredispersión grande	0.2404	0.2443	0.81	0.3252	0.4521
Sobredispersión grande	0.701	0.7048	0.9835	0.7842	0.8667
Sobredispersión grande	0.047	0.0474	0.4221	0.0659	0.1116
Sobredispersión grande	0.0648	0.0651	0.2731	0.0914	0.1359
Sobredispersión grande	0.3433	0.345	0.8698	0.4112	0.5115
Sobredispersión grande	0.8607	0.8593	0.9971	0.8992	0.9354
Sobredispersión grande	0.052	0.0494	0.424	0.0543	0.0594
Sobredispersión grande	0.2047	0.2029	0.6891	0.2128	0.2234
Sobredispersión grande	0.9988	0.9987	1	0.9992	0.9994
Sobredispersión grande	1	1	1	1	1
Sobredispersión grande	0.0518	0.0515	0.4337	0.1033	0.2057
Sobredispersión grande	0.0589	0.0582	0.4611	0.1163	0.2285
Sobredispersión grande	0.1418	0.1525	0.7051	0.2457	0.4268
Sobredispersión grande	0.4253	0.472	0.9481	0.5913	0.7939
Sobredispersión grande	0.0493	0.051	0.4212	0.0795	0.1417
Sobredispersión grande	0.0596	0.0612	0.4726	0.0945	0.1679
Sobredispersión grande	0.22	0.2354	0.7952	0.3082	0.4457
Sobredispersión grande	0.6951	0.7389	0.9891	0.7873	0.8975
Sobredispersión grande	0.0495	0.0501	0.4316	0.0718	0.1149
Sobredispersión grande	0.0668	0.0679	0.4899	0.0934	0.1457
Sobredispersión grande	0.3119	0.3305	0.8511	0.3792	0.4941
Sobredispersión grande	0.8662	0.892	0.9972	0.9073	0.9547
Sobredispersión grande	0.0459	0.0471	0.4249	0.0497	0.0561
Sobredispersión grande	0.2187	0.2223	0.7326	0.2261	0.2422
Sobredispersión grande	0.9985	0.9987	0.9999	0.9986	0.999
Sobredispersión grande	1	1	1	1	1
Sobredispersión mediana	0.0494	0.0496	0.1552	0.1005	0.11
Sobredispersión mediana	0.0516	0.0552	0.1782	0.1078	0.1263
Sobredispersión mediana	0.1115	0.1128	0.3318	0.2018	0.2483
Sobredispersión mediana	0.2876	0.2886	0.6295	0.4257	0.5179
Sobredispersión mediana	0.0498	0.05	0.1576	0.0821	0.0985
Sobredispersión mediana	0.0552	0.0566	0.1771	0.0912	0.1103
Sobredispersión mediana	0.1775	0.1792	0.4236	0.2455	0.2963
Sobredispersión mediana	0.4584	0.4604	0.7747	0.5628	0.6428
Sobredispersión mediana	0.05	0.0505	0.1586	0.0714	0.0878

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.6. Valores de potencia de los modelos utilizados en la investigación.

Datos	Normal	Normal transformado	Poisson	Quasipoisson	Binomial negativo
Sobredispersión mediana	0.0642	0.0658	0.1936	0.0933	0.115
Sobredispersión mediana	0.2285	0.2273	0.4894	0.2854	0.3388
Sobredispersión mediana	0.6163	0.6213	0.8675	0.6897	0.748
Sobredispersión mediana	0.0494	0.0479	0.1593	0.0532	0.0561
Sobredispersión mediana	0.1587	0.1573	0.3413	0.1651	0.1734
Sobredispersión mediana	0.9535	0.9526	0.9889	0.9563	0.9588
Sobredispersión mediana	1	1	1	1	1
Sobredispersión mediana	0.0463	0.0468	0.1526	0.0952	0.1071
Sobredispersión mediana	0.0598	0.0577	0.1869	0.1106	0.1305
Sobredispersión mediana	0.1175	0.1151	0.3341	0.2102	0.2482
Sobredispersión mediana	0.2791	0.2654	0.6078	0.4182	0.4926
Sobredispersión mediana	0.0488	0.0485	0.1557	0.0804	0.097
Sobredispersión mediana	0.0589	0.0571	0.1912	0.0936	0.117
Sobredispersión mediana	0.1791	0.1739	0.4249	0.2437	0.2918
Sobredispersión mediana	0.4498	0.4285	0.7392	0.5424	0.6023
Sobredispersión mediana	0.049	0.0481	0.1558	0.0709	0.089
Sobredispersión mediana	0.0635	0.0635	0.2005	0.0882	0.1096
Sobredispersión mediana	0.2302	0.2208	0.4903	0.2848	0.3315
Sobredispersión mediana	0.5986	0.5784	0.8338	0.6628	0.7077
Sobredispersión mediana	0.0519	0.0518	0.1574	0.0553	0.0586
Sobredispersión mediana	0.1543	0.1538	0.3495	0.1607	0.1672
Sobredispersión mediana	0.9462	0.9413	0.9871	0.9489	0.9498
Sobredispersión mediana	1	1	1	1	1
Sobredispersión mediana	0.0497	0.0476	0.2248	0.0972	0.153
Sobredispersión mediana	0.0521	0.0523	0.231	0.1012	0.164
Sobredispersión mediana	0.153	0.1522	0.5253	0.2524	0.3878
Sobredispersión mediana	0.4361	0.4428	0.8553	0.5801	0.7396
Sobredispersión mediana	0.051	0.0513	0.2327	0.0806	0.1313
Sobredispersión mediana	0.0611	0.0595	0.246	0.0952	0.1461
Sobredispersión mediana	0.247	0.2471	0.6303	0.3244	0.4425
Sobredispersión mediana	0.699	0.6963	0.9562	0.7754	0.8609
Sobredispersión mediana	0.0463	0.0457	0.2273	0.0676	0.1045
Sobredispersión mediana	0.0586	0.0591	0.2529	0.0855	0.1369
Sobredispersión mediana	0.3459	0.3445	0.7248	0.4116	0.5137
Sobredispersión mediana	0.861	0.8598	0.986	0.897	0.9331
Sobredispersión mediana	0.049	0.0484	0.2264	0.0514	0.0566
Sobredispersión mediana	0.1894	0.1866	0.4654	0.1971	0.2092
Sobredispersión mediana	0.9986	0.9986	1	0.9989	0.9989
Sobredispersión mediana	1	1	1	1	1
Sobredispersión mediana	0.0524	0.0527	0.2277	0.098	0.1559
Sobredispersión mediana	0.0589	0.06	0.2519	0.1109	0.1754
Sobredispersión mediana	0.1567	0.1657	0.5225	0.2583	0.3978
Sobredispersión mediana	0.4311	0.4667	0.8612	0.5784	0.7591
Sobredispersión mediana	0.0499	0.0504	0.226	0.0756	0.1262
Sobredispersión mediana	0.0666	0.0676	0.2707	0.0995	0.16
Sobredispersión mediana	0.2378	0.2523	0.6325	0.3137	0.4449
Sobredispersión mediana	0.6972	0.7339	0.962	0.785	0.8868
Sobredispersión mediana	0.0482	0.0475	0.2281	0.0687	0.1045
Sobredispersión mediana	0.0674	0.0687	0.2817	0.0928	0.1387
Sobredispersión mediana	0.3382	0.3521	0.7347	0.4028	0.5177
Sobredispersión mediana	0.8615	0.8797	0.9879	0.8979	0.9473
Sobredispersión mediana	0.0513	0.0523	0.2331	0.0549	0.0601
Sobredispersión mediana	0.2334	0.2357	0.5439	0.2427	0.2611
Sobredispersión mediana	0.9991	0.9991	0.9999	0.9991	0.9995
Sobredispersión mediana	1	1	1	1	1
Sobredispersión pequeña	0.0478	0.0499	0.1561	0.099	0.1106
Sobredispersión pequeña	0.0593	0.0603	0.2841	0.1133	0.159
Sobredispersión pequeña	0.0989	0.0994	0.2956	0.1831	0.2186
Sobredispersión pequeña	0.2249	0.2268	0.5408	0.3548	0.4328
Sobredispersión pequeña	0.0468	0.0478	0.1549	0.0797	0.0932
Sobredispersión pequeña	0.0598	0.0614	0.1836	0.0933	0.1154
Sobredispersión pequeña	0.1483	0.1494	0.3721	0.214	0.2661

(Continúa en la página siguiente.)

(Viene de la página anterior)

Tabla 5.6. Valores de potencia de los modelos utilizados en la investigación.

Datos	Normal	Normal transformado	Poisson	Quasipoisson	Binomial negativo
Sobredispersión pequeña	0.379	0.3826	0.6838	0.4809	0.5501
Sobredispersión pequeña	0.0515	0.0519	0.157	0.0733	0.089
Sobredispersión pequeña	0.0632	0.0635	0.3014	0.0871	0.1169
Sobredispersión pequeña	0.2021	0.202	0.4421	0.2551	0.3013
Sobredispersión pequeña	0.5168	0.5186	0.7884	0.5927	0.653
Sobredispersión pequeña	0.0478	0.0475	0.1568	0.0519	0.0543
Sobredispersión pequeña	0.1366	0.1353	0.3123	0.1426	0.1482
Sobredispersión pequeña	0.9106	0.9086	0.9774	0.9154	0.9202
Sobredispersión pequeña	1	1	1	1	1
Sobredispersión pequeña	0.0482	0.0471	0.1495	0.0982	0.1059
Sobredispersión pequeña	0.0557	0.0555	0.1699	0.1125	0.1223
Sobredispersión pequeña	0.1073	0.1031	0.306	0.1932	0.2258
Sobredispersión pequeña	0.2323	0.2204	0.5186	0.3608	0.4176
Sobredispersión pequeña	0.0528	0.0534	0.1589	0.086	0.1048
Sobredispersión pequeña	0.0538	0.0543	0.1684	0.0874	0.1066
Sobredispersión pequeña	0.149	0.1463	0.3628	0.2116	0.2525
Sobredispersión pequeña	0.3737	0.3531	0.65	0.4698	0.526
Sobredispersión pequeña	0.0476	0.0482	0.1551	0.0706	0.0882
Sobredispersión pequeña	0.0623	0.063	0.1796	0.0866	0.104
Sobredispersión pequeña	0.201	0.1958	0.4303	0.2495	0.2925
Sobredispersión pequeña	0.506	0.4848	0.7561	0.5727	0.616
Sobredispersión pequeña	0.0513	0.0528	0.156	0.0549	0.0584
Sobredispersión pequeña	0.1397	0.1384	0.3059	0.1451	0.1499
Sobredispersión pequeña	0.8934	0.8873	0.9696	0.8987	0.8999
Sobredispersión pequeña	0.9999	0.9999	1	0.9999	0.9999
Sobredispersión pequeña	0.0522	0.0524	0.2294	0.1	0.1581
Sobredispersión pequeña	0.0556	0.054	0.2505	0.1055	0.1705
Sobredispersión pequeña	0.1224	0.124	0.4546	0.211	0.3379
Sobredispersión pequeña	0.3549	0.3563	0.7746	0.4899	0.6575
Sobredispersión pequeña	0.0495	0.0495	0.2249	0.078	0.1264
Sobredispersión pequeña	0.0559	0.0546	0.2563	0.087	0.1411
Sobredispersión pequeña	0.1953	0.1941	0.5454	0.2653	0.3738
Sobredispersión pequeña	0.5778	0.5782	0.8972	0.6708	0.7787
Sobredispersión pequeña	0.0496	0.0484	0.2304	0.0685	0.1099
Sobredispersión pequeña	0.0666	0.0654	0.4549	0.0937	0.1426
Sobredispersión pequeña	0.2662	0.2662	0.6239	0.3242	0.4159
Sobredispersión pequeña	0.7574	0.7581	0.9597	0.8081	0.8693
Sobredispersión pequeña	0.051	0.0514	0.2287	0.054	0.0594
Sobredispersión pequeña	0.1627	0.1623	0.4546	0.1685	0.1802
Sobredispersión pequeña	0.9913	0.9913	0.9989	0.9918	0.9928
Sobredispersión pequeña	1	1	1	1	1
Sobredispersión pequeña	0.0473	0.0462	0.2267	0.0931	0.1522
Sobredispersión pequeña	0.0559	0.0569	0.2374	0.1082	0.1676
Sobredispersión pequeña	0.1253	0.1313	0.4539	0.2148	0.3426
Sobredispersión pequeña	0.3344	0.3648	0.7972	0.4863	0.6764
Sobredispersión pequeña	0.0486	0.0495	0.2239	0.0779	0.1256
Sobredispersión pequeña	0.0585	0.0587	0.2443	0.092	0.1448
Sobredispersión pequeña	0.19	0.1974	0.5487	0.2594	0.3742
Sobredispersión pequeña	0.577	0.6031	0.9157	0.6674	0.7916
Sobredispersión pequeña	0.0478	0.0484	0.2233	0.0679	0.1067
Sobredispersión pequeña	0.065	0.0652	0.2507	0.0911	0.13
Sobredispersión pequeña	0.2653	0.2749	0.6336	0.3239	0.4252
Sobredispersión pequeña	0.7476	0.7692	0.9655	0.8016	0.8799
Sobredispersión pequeña	0.0511	0.051	0.2298	0.0538	0.0597
Sobredispersión pequeña	0.1698	0.1689	0.4479	0.1767	0.1892
Sobredispersión pequeña	0.9914	0.9917	0.9993	0.9921	0.9933
Sobredispersión pequeña	1	1	1	1	1

(Fin de la tabla.)

Listing 5.1. Código utilizado para la simulación Monte Carlo.

```
# Programas que utiliza la simulacion -----
```

```

library(MASS)
library(AER) # Para probar sobredispersión <- dispersiontest(modelo)
# Elementos que van dentro de la función -----
trat <- c(4, 4, 8, 8)
rep <- c(4, 6, 8, 50)
sobredis <- c(20, 40, 100) # 1.32 ejem Agresti pag. 255 agregar a la simulación
tam_efec <- c(0, .01, .1, .25)
simula <- 10000
# Medias para los 4 casos -----
media01 <- matrix(c(20,20,20,20,
                  18,18,18.7,18.7,
                  20.2,20.2,22.9,22.9,
                  20.4,20.4,25.4,25.4), ncol =4, byrow = TRUE)
media02 <- matrix(c(20,20,20,20,
                  23.7,22.7,22.7,22.7,
                  21.7,18.6,18.6,18.6,
                  27.4,21.6,21.6,21.6), ncol =4, byrow = TRUE)
media03 <- matrix(c(20,20,20,20,20,20,20,20,
                  18,18,18,18,18.7,18.7,18.7,18.7,
                  20.2,20.2,20.2,20.2,22.9,22.9,22.9,22.9,
                  20.4,20.4,20.4,20.4,25.4,25.4,25.4,25.4), ncol =8, byrow = TRUE)
media04 <- matrix(c(20,20,20,20,20,20,20,20,
                  19.4,19.4,20.3,20.3,20.3,20.3,20.3,20.3,
                  17.1,17.1,20.1,20.1,20.1,20.1,20.1,20.1,
                  17.3,17.3,22.8,22.8,22.8,22.8,22.8,22.8), ncol =8, byrow = TRUE)
media05 <- matrix(c(20,20,20,20,
                  19.1,19.1,20.2,20.2,
                  20.5,20.5,24.7,24.7,
                  19.3,19.3,26.6,26.6), ncol = 4, byrow = TRUE)
media06 <- matrix(c(20,20,20,20,
                  23.1,21.7,21.7,21.7,
                  27.1,22.2,22.2,22.2,
                  29.8,21.3,21.3,21.3), ncol = 4, byrow = TRUE)
media07 <- matrix(c(20,20,20,20,20,20,20,20,
                  18.3,18.3,18.3,18.3,19.4,19.4,19.4,19.4,
                  20.5,20.5,20.5,20.5,24.8,24.8,24.8,24.8,
                  19.2,19.2,19.2,19.2,26.8,26.8,26.8,26.8), ncol = 8, byrow = TRUE)
media08 <- matrix(c(20,20,20,20,20,20,20,20,
                  19.2,19.2,20.6,20.6,20.6,20.6,20.6,20.6,
                  17.8,17.8,22.4,22.4,22.4,22.4,22.4,22.4,
                  15,15,23,23,23,23,23,23), ncol = 8, byrow = TRUE)
media09 <- matrix(c(20,20,20,20,
                  19.2,19.2,20.2,20.2,
                  20.4,20.4,23.9,23.9,
                  19.8,19.8,26,26), ncol = 4, byrow = TRUE)
media10 <- matrix(c(20,20,20,20,
                  23.1,21.9,21.9,21.9,
                  26.5,22.3,22.3,22.3,
                  28.9,21.6,21.6,21.6), ncol = 4, byrow = TRUE)
media11 <- matrix(c(20,20,20,20,20,20,20,20,
                  18.4,18.4,18.4,18.4,19.3,19.3,19.3,19.3,
                  20.7,20.7,20.7,20.7,24.4,24.4,24.4,24.4,
                  19.7,19.7,19.7,19.7,26.2,26.2,26.2,26.2), ncol = 8, byrow = TRUE)
media12 <- matrix(c(20,20,20,20,20,20,20,20,
                  19.1,19.1,20.3,20.3,20.3,20.3,20.3,20.3,
                  18.3,18.3,22.4,22.4,22.4,22.4,22.4,22.4,
                  15.9,15.9,22.8,22.8,22.8,22.8,22.8,22.8), ncol = 8, byrow = TRUE)
media13 <- matrix(c(20,20,20,20,
                  18,18,18.9,18.9,
                  19.5,19.5,22.8,22.8,
                  18.5,18.5,24.2,24.2), ncol = 4, byrow = TRUE)
media14 <- matrix(c(20,20,20,20,
                  20.5,19.4,19.4,19.4,
                  24.1,20.3,20.3,20.3,
                  25.1,18.7,18.7,18.7), ncol = 4, byrow = TRUE)
media15 <- matrix(c(20,20,20,20,20,20,20,20,
                  18,18,18,18,19,19,19,19,
                  19.5,19.5,19.5,19.5,22.9,22.9,22.9,22.9,

```

```

      18.4,18.4,18.4,18.4,24.4,24.4,24.4,24.4), ncol = 8, byrow = TRUE)
media16 <- matrix(c(20,20,20,20,20,20,20,20,
      18.3,18.3,19.4,19.4,19.4,19.4,19.4,19.4,
      18.1,18.1,21.9,21.9,21.9,21.9,21.9,21.9,
      17.1,17.1,24.1,24.1,24.1,24.1,24.1,24.1), ncol = 8, byrow = TRUE)
media17 <- matrix(c(20,20,20,20,
      20.3,20.3,21.2,21.2,
      19.6,19.6,22.7,22.7,
      18.7,18.7,24,24), ncol = 4, byrow = TRUE)
media18 <- matrix(c(20,20,20,20,
      20.2,19.2,19.2,19.2,
      24.2,20.6,20.6,20.6,
      24.9,18.9,18.9,18.9), ncol = 4, byrow = TRUE)
media19 <- matrix(c(20,20,20,20,20,20,20,20,
      20.3,20.3,20.3,20.3,21.2,21.2,21.2,21.2,
      19.6,19.6,19.6,19.6,22.7,22.7,22.7,22.7,
      18.6,18.6,18.6,18.6,24.1,24.1,24.1,24.1), ncol = 8, byrow = TRUE)
media20 <- matrix(c(20,20,20,20,20,20,20,20,
      18.1,18.1,19.1,19.1,19.1,19.1,19.1,19.1,
      18,18,21.5,21.5,21.5,21.5,21.5,21.5,
      17.4,17.4,23.8,23.8,23.8,23.8,23.8,23.8), ncol = 8, byrow = TRUE)
# Generador de datos: Poisson -----
dist_pois <- function(trat, rep, media){
  output <- vector("numeric",3)
  for (i in 1:trat) {
    for (j in 1:rep) {
      y <- rpois(n = 1, lambda = media[i])
      x <- c(i, j, y)
      output <- rbind(output, x)
    }
  }
  a <- trat*rep + 1
  output <- output[2:a,]
}
# Simulacion de los modelos: Poisson -----
output <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
potencia_f <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
for (i in 1:4) {
  # numero_tratamientos
  if(i == 1) {media=media01}
  if(i == 2) {media=media02}
  if(i == 3) {media=media03}
  if(i == 4) {media=media04}
  for (j in 1:4) {
    # numero_repeticiones
    for (k in 1:4) {
      # tamano_del_efecto
      potencia <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
      for (a in 1:simula) {
        # numero_de_simulacion
        data <- dist_pois(trat[i],rep[j],media[k,])
        modelo_norm <- anova(lm( data[,3] ~ factor(data[,1])))
        vpMN <- modelo_norm[1,5]
        if (vpMN < .05) { potencia[5] <- potencia[5] + 1}
        modelo_transf <- anova(lm( sqrt(data[,3]) ~ factor(data[,1])))
        vpMT <- modelo_transf[1,5]
        if (vpMT < .05) { potencia[6] <- potencia[6] + 1}
        modelo_pois <- anova(glm( data[,3] ~ factor(data[,1]),
          family = poisson(link = log)))
        vpMP <- pchisq (modelo_pois$Deviance[2],trat[i]-1,lower.tail = F)
        if (vpMP < .05) { potencia[7] <- potencia[7] + 1}
        prueba <- dispersiontest(glm( data[,3] ~ factor(data[,1]),
          family = poisson(link = "log")))
        vpSD <- prueba[2]
        if (vpSD < .05) {potencia[10] <- potencia[10] + 1}
        modelo_qpois <- anova(glm( data[,3] ~ factor(data[,1]),
          family = quasipoisson(link = "log")), test = "Rao")
        vpMQP <- modelo_qpois[2,6]
        if (is.na(vpMQP)) {vpMQP = 1}
        if (vpMQP < .05) { potencia[8] <- potencia[8] + 1}
        modelo_bn <- anova(MASS::glm.nb( data[,3] ~ factor(data[,1]),
          init.theta = 1, link = log))
      }
    }
  }
}

```

```

vpMBN <- pchisq (modelo_bn$Deviance[2],trat[i]-1,lower.tail = F)
if (vpMBN < .05) { potencia[9] <- potencia[9] + 1}
output2 <- c(i, trat[i], rep[j], tam_efec[k], vpMN, vpMT, vpMP, vpMQP, vpMBN)
output <- rbind(output, output2)
}
potencia2 <- c(i, trat[i], rep[j], tam_efec[k],
potencia[5], potencia[6], potencia[7], potencia[8], potencia[9],
potencia [10])
potencia_f <- rbind(potencia_f, potencia2)
}
}
# fin_de_la_simulacion
dimnames(output) <- list(NULL, c("Caso", "Tratamiento","Repeticion","T_Efecto","Normal", "↔
Transformado", "Poisson", "Quasipoisson", "B_Negativo", "Contador_sobredispersión"))
output <- output[2:(a*i*k*j+1),]
dimnames(potencia_f) <- list(NULL, c("Caso", "Tratamiento","Repeticion","T_Efecto","Normal↔
", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "Contador_sobredispersión"↔
))
potencia_f <- potencia_f[2:(i*k*j+1),]
potencia <- cbind(potencia_f[,1:4], potencia_f[,5:10]/simula)
# Generador de datos: sobredispersión -----
dist_bn <- function(trat, rep, sobredis, media){
output <- vector("numeric",3)
for (i in 1:trat) {
for (j in 1:rep) {
y <- rbinom(n = 1, size = sobredis, mu = media[i])
x <- c(i, j, y)
output <- rbind(output, x)
}
}
}
a <- trat*rep + 1
output <- output[2:a,]
}
# Simulación de los modelos: sobredispersión 20 -----
output <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
potencia_f <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
for (i in 1:4) { # numero_tratamientos
if(i == 1) {media = media05}
if(i == 2) {media = media06}
if(i == 3) {media = media07}
if(i == 4) {media = media08}
for (j in 1:4) { # numero_repeticiones
for (k in 1:1) { # sobredispersión
for (l in 1:4) { # tamaño_del_efecto
potencia <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
for (a in 1:simula) { # número_de_simulación
data <- dist_bn(trat[i],rep[j], sobredis[k], media[l,])
modelo_norm <- anova(lm( data[,3] ~ factor(data[,1])))
vpMN <- modelo_norm[1,5]
if (vpMN < .05) { potencia[6] <- potencia[6] + 1}
modelo_transf <- anova(lm( sqrt(data[,3]) ~ factor(data[,1])))
vpMT <- modelo_transf[1,5]
if (vpMT < .05) { potencia[7] <- potencia[7] + 1}
modelo_pois <- anova(glm( data[,3] ~ factor(data[,1]),
family = poisson(link = "log")))
vpMP <- pchisq (modelo_pois$Deviance[2],trat[i]-1,lower.tail = F)
if (vpMP < .05) { potencia[8] <- potencia[8] + 1}
prueba <- dispersionstest(glm( data[,3] ~ factor(data[,1]),
family = poisson(link = "log")))
vpSD <- prueba[2]
if (vpSD < .05) {potencia[11] <- potencia[11] + 1}
modelo_qpois <- anova(glm( data[,3] ~ factor(data[,1]),
family = quasipoisson(link = "log")), test = "Rao")
vpMQP <- modelo_qpois [2,6]
if (is.na(vpMQP)) {vpMQP = 1}
if (vpMQP < .05) { potencia[9] <- potencia[9] + 1}
modelo_bn <- anova(MASS::glm.nb( data[,3] ~ factor(data[,1]),
init.theta = sobredis, link = log))

```

```

vpMBN <- pchisq (modelo_bn$Deviance[2],trat[i]-1,lower.tail = F)
if (vpMBN < .05) { potencia[10] <- potencia[10] + 1}
output2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[1],
            vpMN, vpMT, vpMP, vpMQP, vpMBN, vpSD)
output <- rbind(output, output2)
}
potencia2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[1],
            potencia[6], potencia[7], potencia[8], potencia[9], potencia[10],
            potencia[11])
potencia_f <- rbind(potencia_f, potencia2)
}
}
}
}
# fin_de_la_simulacion
dimnames(output) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion", "←
T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "←
Contador_sobredispersion"))
output <- output[2:(a*i*k*j*1+1),]
dimnames(potencia_f) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion"←
, "T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "←
Contador_sobredispersion"))
potencia_f <- potencia_f[2:(i*k*j*1+1),]
potencia <- cbind(potencia_f[,1:5], potencia_f[,6:11]/simula)
# Simulacion de los modelos: sobredispersion 40 -----
sobredis <- c(40)
output <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
potencia_f <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
for (i in 1:4) { # numero_tratamientos
  if(i == 1) {media = media09}
  if(i == 2) {media = media10}
  if(i == 3) {media = media11}
  if(i == 4) {media = media12}
  for (j in 1:4) { # numero_repeticiones
    for (k in 1:1) { # sobredispersion
      for (l in 1:4) { # tamaño_del_efecto
        potencia <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
        for (a in 1:simula) { # numero_de_simulacion
          data <- dist_bn(trat[i],rep[j], sobredis[k], media[l,])
          modelo_norm <- anova(lm( data[,3] ~ factor(data[,1])))
          vpMN <- modelo_norm[1,5]
          if (vpMN < .05) { potencia[6] <- potencia[6] + 1}
          modelo_transf <- anova(lm( sqrt(data[,3]) ~ factor(data[,1])))
          vpMT <- modelo_transf[1,5]
          if (vpMT < .05) { potencia[7] <- potencia[7] + 1}
          modelo_pois <- anova(glm( data[,3] ~ factor(data[,1]),
                                family = poisson(link = "log")))
          vpMP <- pchisq (modelo_pois$Deviance[2],trat[i]-1,lower.tail = F)
          if (vpMP < .05) { potencia[8] <- potencia[8] + 1}
          prueba <- dispersiontest(glm( data[,3] ~ factor(data[,1]),
                                       family = poisson(link = "log")))
          vpSD <- prueba[2]
          if (vpSD < .05) {potencia[11] <- potencia[11] + 1}
          modelo_qpois <- anova(glm( data[,3] ~ factor(data[,1]),
                                   family = quasipoisson(link = "log")), test = "Rao")
          vpMQP <- modelo_qpois[2,6]
          if (is.na(vpMQP)) {vpMQP = 1}
          if (vpMQP < .05) { potencia[9] <- potencia[9] + 1}
          modelo_bn <- anova(MASS::glm.nb( data[,3] ~ factor(data[,1]),
                                          init.theta = sobredis, link = log))
          vpMBN <- pchisq (modelo_bn$Deviance[2],trat[i]-1,lower.tail = F)
          if (vpMBN < .05) { potencia[10] <- potencia[10] + 1}
          output2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[1],
                    vpMN, vpMT, vpMP, vpMQP, vpMBN, vpSD)
          output <- rbind(output, output2)
        }
      }
    }
  }
  potencia2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[1],
                potencia[6], potencia[7], potencia[8], potencia[9],
                potencia[10], potencia[11])
}

```

```

        potencia_f <- rbind(potencia_f, potencia2)
    }
}
}
# fin_de_la_simulacion
dimnames(output) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion", "←
T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "←
Contador_sobredispersion"))
output <- output[2:(a*i*k*j*l+1),]
dimnames(potencia_f) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion"←
, "T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "←
Contador_sobredispersion"))
potencia_f <- potencia_f[2:(i*k*j*l+1),]
potencia <- cbind(potencia_f[,1:5], potencia_f[,6:11]/simula)
# Simulacion de los modelos: sobredispersion 100 -----
sobredis <- c(100)
output <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
potencia_f <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
for (i in 1:4) {
    # numero_tratamientos
    if(i == 1) {media = media17}
    if(i == 2) {media = media18}
    if(i == 3) {media = media19}
    if(i == 4) {media = media20}
    for (j in 1:4) {
        # numero_repeticiones
        for (k in 1:1) {
            # sobredispersion
            for (l in 1:4) {
                # tamano_del_efecto
                potencia <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
                for (a in 1:simula) {
                    # numero_de_simulacion
                    data <- dist_bn(trat[i], rep[j], sobredis[k], media[l,])
                    modelo_norm <- anova(lm( data[,3] ~ factor(data[,1])))
                    vpMN <- modelo_norm[1,5]
                    if (vpMN < .05) { potencia[6] <- potencia[6] + 1}
                    modelo_transf <- anova(lm( sqrt(data[,3]) ~ factor(data[,1])))
                    vpMT <- modelo_transf[1,5]
                    if (vpMT < .05) { potencia[7] <- potencia[7] + 1}
                    modelo_pois <- anova(glm( data[,3] ~ factor(data[,1]),
                                            family = poisson(link = "log")))
                    vpMP <- pchisq (modelo_pois$Deviance[2], trat[i]-1, lower.tail = F)
                    if (vpMP < .05) { potencia[8] <- potencia[8] + 1}
                    prueba <- dispersionstest(glm( data[,3] ~ factor(data[,1]),
                                                  family = poisson(link = "log")))
                    vpSD <- prueba[2]
                    if (vpSD < .05) {potencia[11] <- potencia[11] + 1}
                    modelo_qpois <- anova(glm( data[,3] ~ factor(data[,1]),
                                              family = quasipoisson(link = "log"), test = "Rao")
                    vpMQP <- modelo_qpois[2,6]
                    if (is.na(vpMQP)) {vpMQP = 1}
                    if (vpMQP < .05) { potencia[9] <- potencia[9] + 1}
                    modelo_bn <- anova(MASS::glm.nb( data[,3] ~ factor(data[,1]),
                                                    init.theta = sobredis, link = log))
                    vpMBN <- pchisq (modelo_bn$Deviance[2], trat[i]-1, lower.tail = F)
                    if (vpMBN < .05) { potencia[10] <- potencia[10] + 1}
                    output2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[l],
                                vpMN, vpMT, vpMP, vpMQP, vpMBN, vpSD)
                    output <- rbind(output, output2)
                }
            }
        }
    }
    potencia2 <- c(i, trat[i], rep[j], sobredis[k], tam_efec[l],
                potencia[6], potencia[7], potencia[8], potencia[9],
                potencia[10], potencia[11])
    potencia_f <- rbind(potencia_f, potencia2)
}
}
}
# fin_de_la_simulacion
dimnames(output) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion",
"T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "←
Contador_sobredispersion"))
output <- output[2:(a*i*k*j*l+1),]

```

```

dimnames(potencia_f) <- list(NULL, c("Caso", "Tratamiento", "Repeticion", "Sobredispersion"↵
,
"T_Efecto", "Normal", "Transformado", "Poisson", "Quasipoisson", "B_Negativo", "↵
Contador_sobredispersion"))
potencia_f <- potencia_f[2:(i*k*j*1+1),]
potencia <- cbind(potencia_f[,1:5], potencia_f[,6:11]/simula)
# Tablas de excel -----
setwd("/Users/arwell/Desktop/Simulaciones")
write.table(potencia, file = "poisson_10000.csv", sep = ";", row.names = FALSE)
write.table(output, file = "output_poisson_10000.csv", sep = ";", row.names = FALSE)
setwd("/Users/arwell/Desktop/Simulaciones")
write.table(potencia, file = "sobredispersion20_10000.csv", sep = ";", row.names = FALSE)
write.table(output, file = "output_sobredispersion20_10000.csv", sep = ";", row.names = ↵
FALSE)
setwd("/Users/arwell/Desktop/Simulaciones")
write.table(potencia, file = "sobredispersion40_10000.csv", sep = ";", row.names = FALSE)
write.table(output, file = "output_sobredispersion40_10000.csv", sep = ";", row.names = ↵
FALSE)
setwd("/Users/arwell/Desktop/Simulaciones")
write.table(potencia, file = "sobredispersion100_10000.csv", sep = ";", row.names = FALSE)
write.table(output, file = "output_sobredispersion100_10000.csv", sep = ";", row.names = ↵
FALSE)

```