

5

Estudios de simulación

El modelado a través de MLG está disponible en muchos de los paquetes estadísticos más conocidos, como Stata, SAS, Limdep, S o R. De estos paquetes quizás sea Limdep (versión 7.0) uno de los más exhaustivos en cuanto a la incorporación de procedimientos de análisis de datos para variables de recuento. Sin embargo, Limdep no presenta la flexibilidad ni la extensibilidad de un entorno para el modelado como R o S. En cuanto a las diferencias entre S y R, este último es «*freeware*», consume menos recursos del sistema, presenta una mayor modularidad que, junto con los continuos desarrollos y aportaciones desinteresadas por parte de especialistas en cada uno de los ámbitos de la estadística, lo hacen más extensible y de una forma más inmediata. Estos han sido algunos de los motivos por los que se ha seleccionado el entorno R para llevar a cabo los estudios que se presentan a continuación.

5.1 Estudio de la tasa nominal de error de los tests diagnósticos de sobredispersión

El objetivo de este experimento de simulación es comparar la tasa nominal de error de los siguientes tests estadísticos utilizados habitualmente en la literatura para valorar la presencia de sobredispersión en los datos de recuento:

- Razón de verosimilitud: $LR \sim \chi^2_{1,2}$
- Test de Wald: $W \sim \chi^2_{1,2}$
- Multiplicador de Lagrange basado en regresión (Negbin I): $LMR I \sim t_{n-1}$
- Multiplicador de Lagrange basado en regresión (Negbin II): $LMR II \sim t_{n-1}$
- χ^2 dividido por grados de libertad: χ^2_{n-p}
- Discrepancia dividido por grados de libertad: $D \sim \chi^2_{n-p}$

Es importante tener en cuenta que, tal como indican diversos autores (Cameron y Trivedi, 1998; Long, 1997), las pruebas LR y W, se implementan como pruebas unilaterales debido a la restricción que no puede ser negativa. De esta forma

los valores críticos de significación quedan establecidos a partir del criterio 2 en lugar de .

Puesto que el objetivo es estudiar el error de primera especie de estas pruebas, se ha procedido a la extracción aleatoria de 5,000 muestras de tamaños:

$$n = 500, n = 100, n = 50 \text{ y } n = 20$$

a partir de distribuciones de Poisson con parámetros :

$$= 0.3, = 1, = 5 \text{ y } = 10$$

De este modo se genera la distribución muestral de estas pruebas bajo la hipótesis de equidispersión (ausencia de sobredispersión) y, por consiguiente, la proporción de decisiones de significación estadística basadas en estas pruebas (en este caso con un nivel $=0.05$) es la estimación empírica de la tasa nominal de error de las mismas. Las siguientes figuras muestran las distribuciones muestrales empíricas obtenidas para las pruebas LR y W.

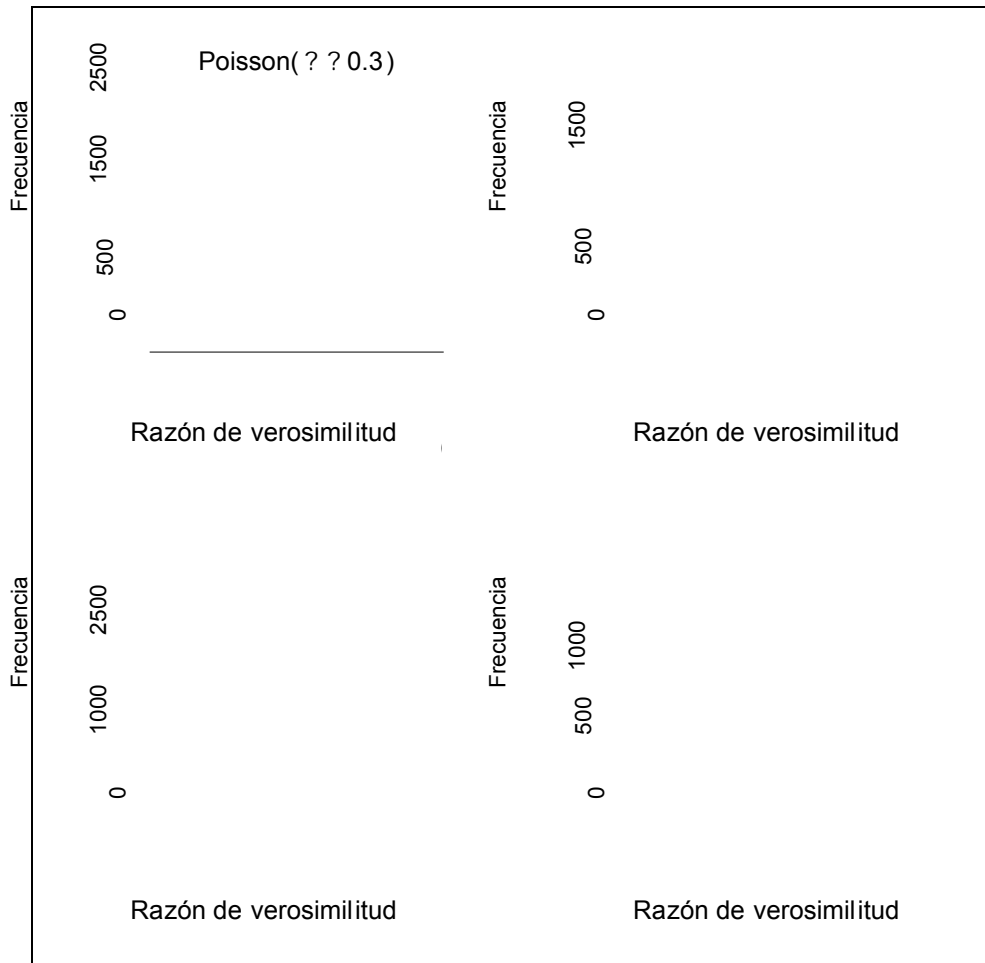


Ilustración 1. Distribuciones muestrales Monte Carlo de la prueba LR para el diagnóstico de la sobredispersión: 5,000 muestras Poisson() de tamaño $n = 500$

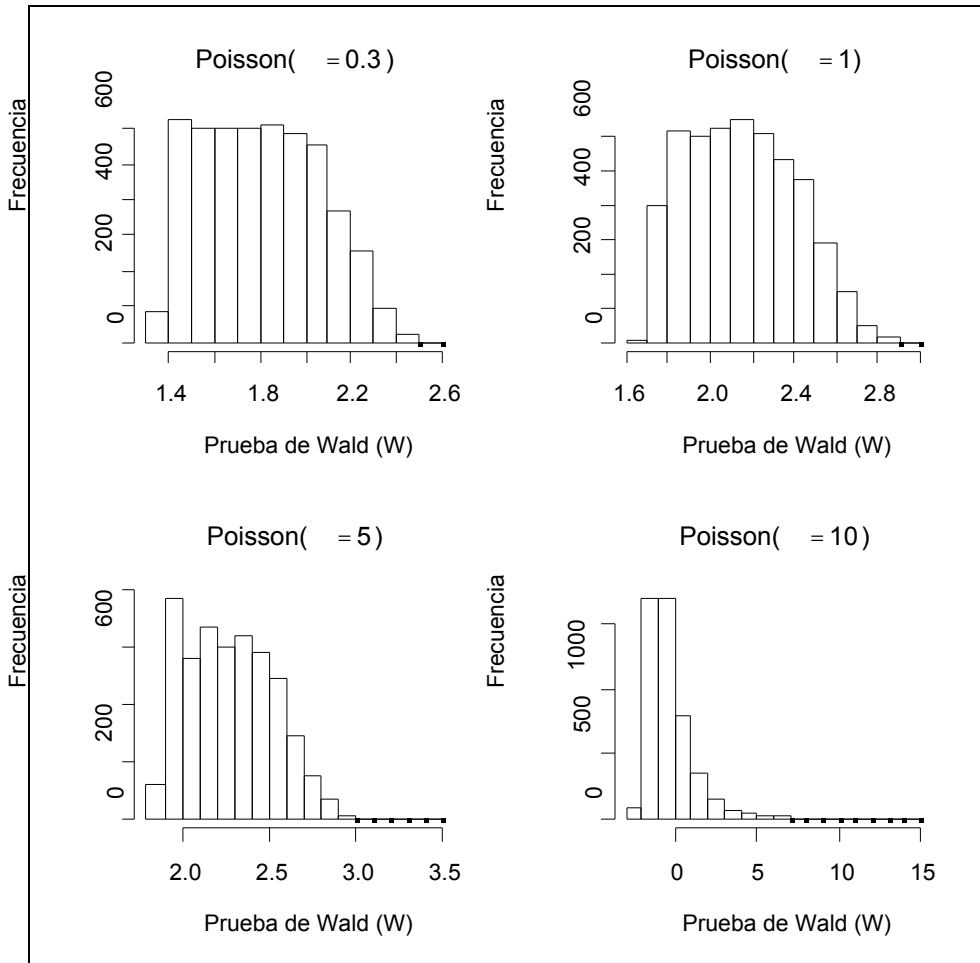


Ilustración 2. Distribuciones muestrales Monte Carlo de la prueba de Wald para el diagnóstico de la sobredispersión: 5,000 muestras Poisson(λ) de tamaño $n = 500$

La Tabla 5 muestra los resultados de este primer experimento de simulación Monte Carlo.

En primer lugar, se observa que, en general, la tasa nominal de error es independiente de λ . Sin embargo, sí es importante el efecto de los tamaños muestrales:

- Para tamaños muestrales elevados, la tasa nominal de error se ajusta al criterio de significación establecido, mientras que ocurre lo contrario en los

tamaños muestrales más bajos, especialmente $n = 20$, mostrando, por tanto, el fuerte carácter asintótico de la validez de estas pruebas.

Tabla 5. Muestras Poisson:
tasa nominal de error de los tests de sobredispersión.

n	λ	LR			Wald			LMR Negbin I		
		LR	P _{2.5}	P _{97.5}	W	P _{2.5}	P _{97.5}	LMR I	P _{2.5}	P _{97.5}
20	0.3	0.012	-0.097	1.991	0.000	0.424	0.887	0.006	-1.263	1.295
20	1	0.019	-0.113	2.260	0.000	0.490	1.018	0.093	-1.368	1.370
20	5	0.023	-0.115	2.580	0.000	0.509	1.159	0.166	-1.526	1.370
20	10	0.021	-0.115	2.474	0.000	0.507	1.156	0.171	-1.569	1.339
50	0.3	0.023	-0.187	2.622	0.000	0.593	1.106	0.080	-1.381	1.379
50	1	0.031	-0.221	3.001	0.000	0.683	1.291	0.090	-1.151	1.519
50	5	0.027	-0.232	2.810	0.000	0.707	1.341	0.100	-1.095	1.443
50	10	0.025	-0.235	2.699	0.000	0.717	1.328	0.105	-1.124	1.471
100	0.3	0.031	-0.300	3.017	0.000	0.764	1.303	0.091	-1.168	1.463
100	1	0.035	-0.357	3.265	0.000	0.893	1.457	0.078	-0.976	1.555
100	5	0.030	-0.383	2.974	0.000	0.951	1.465	0.081	-0.964	1.509
100	10	0.036	-0.388	3.213	0.000	0.959	1.511	0.084	-0.963	1.608
500	0.3	0.041	-0.920	3.530	0.000	1.589	2.297	0.066	-0.886	1.625
500	1	0.040	-1.157	3.376	0.012	1.939	2.643	0.053	-0.803	1.718
500	5	0.044	-1.238	3.608	0.045	2.065	2.764	0.055	-0.785	1.740
500	10	0.040	-1.250	3.421	0.046	2.076	2.758	0.051	-0.790	1.728
n	λ	LMR Negbin II			Discrepancia			χ^2		
		LMR II	P _{2.5}	P _{97.5}	D	P _{2.5}	P _{97.5}	χ^2	P _{2.5}	P _{97.5}
20	0.3	0.025	-1.353	1.399	0.001	12.673	22.696	0.040	14.344	31.058
20	1	0.092	-1.374	1.373	0.065	17.503	31.629	0.043	14.034	31.285
20	5	0.166	-1.518	1.360	0.078	14.376	34.052	0.049	13.849	31.885
20	10	0.172	-1.569	1.345	0.059	13.946	32.324	0.049	13.735	31.220
50	0.3	0.093	-1.347	1.425	0.000	36.072	52.027	0.048	41.301	69.857
50	1	0.092	-1.154	1.519	0.116	49.881	72.550	0.050	41.405	69.638
50	5	0.010	-1.093	1.445	0.080	43.148	72.533	0.050	41.360	69.091
50	10	0.104	-1.117	1.468	0.063	41.837	69.943	0.049	41.094	68.360
100	0.3	0.089	-1.175	1.473	0.000	76.180	98.926	0.054	88.163	128.536
100	1	0.078	-0.977	1.575	0.206	105.116	136.660	0.053	88.806	128.112
100	5	0.082	-0.961	1.501	0.094	92.229	132.275	0.043	88.402	126.833
100	10	0.083	-0.959	1.600	0.067	89.792	129.943	0.053	88.252	128.041
500	0.3	0.069	-0.891	1.649	0.000	405.71	454.96	0.055	476.23	564.829
500	1	0.051	-0.799	1.718	0.780	553.430	623.203	0.050	476.56	561.187
500	5	0.055	-0.788	1.750	0.195	498.947	589.824	0.051	476.528	562.633
500	10	0.051	-0.792	1.725	0.091	485.424	572.691	0.049	476.410	561.190

- Mientras que la prueba LR se muestra como la más eficiente, puesto que la convergencia de las proporciones de significación al criterio establecido es progresiva y constante, en la prueba de Wald ocurre lo contrario: infraestima sistemáticamente la sobredispersión para todos los tamaños muestrales excepto para $n = 500$, donde se acerca a la proporción 0.05. Por su parte, las pruebas LMR Negbin I, LMR Negbin II y Discrepancia se muestran algo más erráticas.

En general, la prueba que muestra más estabilidad bajo las diferentes condiciones experimentales (tamaño muestral y parámetro de localización de la distribución de Poisson) es la basada en el valor χ^2/gl .

5.2 Estudio de diferentes aspectos relacionados con el diagnóstico de la sobredispersión y su modelado

En este segundo grupo de experimentos de simulación el objetivo es estudiar los siguientes aspectos relacionados con la presencia de sobredispersión en el modelado de respuestas de recuento:

- Estimación empírica de la potencia estadística de los tests para el diagnóstico de la sobredispersión presentados en el apartado anterior, definida dicha potencia como la proporción de diagnósticos estadísticamente significativos siendo cierta la presencia de sobredispersión (se utilizará el criterio de significación $\alpha = 0.05$).
- Evaluación de la consistencia y de la eficiencia de las dos estimaciones del parámetro de dispersión más habituales en la práctica, a saber:
 - La estimación del parámetro de dispersión en la distribución Negbin II.
 - La estimación del parámetro de escala mediante quasi-Poisson.
- Evaluación de la consistencia y eficiencia de los coeficientes de regresión estimados mediante MRP, MR Negbin II y MRQP (quasi-Poisson).
- Evaluación de la precisión de las estimaciones de los errores estándar de los coeficientes de los modelos de regresión ajustados.

En estos experimentos, el aspecto clave es la definición de mecanismos estocásticos que generen muestras de datos con diferentes grados (conocidos) de sobredispersión. Concretamente, se han escogido los tres mecanismos siguientes por aparecer como los más habituales en la práctica (Cameron y Trivedi, 1998; Winkelmann, 2000):

- Distribución binomial negativa.
- Mezcla de dos distribuciones de Poisson con diferente parámetro de localización .
- Exceso de ceros, mediante la mezcla de dos mecanismo generadores: un modelo logit de decisión ($Y = 0 / Y \sim \text{Poisson}(\lambda)$), y un modelo de Poisson para el segundo caso.

Para las diferentes muestras generadas bajo cada uno de los mecanismos estocásticos indicados y, como en el experimento anterior, para los tamaños muestrales:

$$n = 500, n = 100, n = 50 \text{ y } n = 20$$

se calculan las mismas pruebas diagnósticas para la sobredispersión estudiadas en el primer experimento de simulación ($LR_{1,2}^2$, $W_{1,2}^2$, $LMR I_{t_{n,1}}$, $LMR II_{t_{n,1}}$, $\chi^2_{n,p}$ y $D_{n,p}^2$), y se ajustan los modelos de regresión simple MRP, MR Negbin II y MRQP. Para cada muestra, la variable de respuesta Y será una variable de recuento distribuida según la ley o mecanismo de probabilidad correspondiente, y la única variable explicativa X no guardará ninguna relación con Y , de modo que es fácil conocer *a priori* los valores que deberán tomar los dos coeficientes b_0 y b_1 estimados mediante los diferentes modelos de regresión:

$$b_0 = \log(\lambda) \rightarrow \exp(b_0) = \lambda$$

$$b_1 = 0 \rightarrow \exp(b_1) = 1$$

5.2.1 Presencia de sobredispersión simulada mediante un modelo Negbin II

Se ha procedido a la extracción aleatoria de 5,000 muestras de tamaños:

$$n = 500, n = 100, n = 50 \text{ y } n = 20$$

a partir de distribuciones de Poisson con parámetros λ :

$$\lambda = 0.3, \lambda = 1 \text{ y } \lambda = 5$$

Para la simulación de sobredispersión a través de la distribución binomial negativa, concretamente con la función variancia correspondiente a Negbin II, con las siguientes configuraciones de los parámetros (se indica la relación entre el valor de los parámetros y los valores esperados para el coeficiente b_0 en los modelos de regresión que se ajustarán en cada muestra, así como el valor esperado del parámetro de escala λ que se estimará mediante quasi-Poisson):

λ	τ	Var = $\lambda + \lambda^2 / \tau$	Valor esperado del parámetro de escala ϕ
0.3	1.2	$0.3 + 0.3^2 / 1.2 = 0.375$	$0.375 / 0.3 = 1.25$
	0.3	$0.3 + 0.3^2 / 0.3 = 0.6$	$0.6 / 0.3 = 2$
	0.15	$0.3 + 0.3^2 / 0.15 = 0.9$	$0.9 / 0.3 = 3$
1	4	$1 + 1^2 / 4 = 1.25$	$1.25 / 1 = 1.25$
	1	$1 + 1^2 / 1 = 2$	$2 / 1 = 2$
	0.5	$1 + 1^2 / 0.5 = 3$	$3 / 1 = 3$
5	20	$5 + 5^2 / 20 = 6.25$	$6.25 / 5 = 1.25$
	5	$5 + 5^2 / 5 = 10$	$10 / 5 = 2$
	2.5	$5 + 5^2 / 2.5 = 15$	$15 / 5 = 3$

1. *Potencia estadística de las pruebas diagnósticas de sobredispersión:* se observa que, en general, todas las pruebas funcionan razonablemente bien para los diferentes niveles de sobredispersión simulados. Se observa que, en general, la potencia mejora a medida que aumenta la sobredispersión y el tamaño muestral (ver Tabla 6 y Tabla 7).
2. *Estimaciones de los parámetros de dispersión:* Con tamaño muestral $n = 500$ las estimaciones de los parámetros de dispersión son casi idénticas a las esperadas, tanto de Negbin II como mediante quasi-Poisson (ver Tabla 8). Para el resto de tamaños muestrales n_i se aproximan, en general, al valor esperado.
3. *Estabilidad de las estimaciones de β_0 y β_1 a lo largo de los diferentes modelos de regresión ajustados:* Tal como indican entre otros Gourieroux, Monfort y Trognon (1984a) las estimaciones del MRP de los coeficientes de regresión resultan generalmente insesgadas aún en presencia de sobredispersión. En efecto, tal como se muestra en la Tabla 9 y la Tabla 10, los valores de β_0 y β_1 obtenidos mediante MRP, MRQP y MR Negbin II son muy parecidos. La convergencia mutua de los valores de los parámetros se hace más evidente a medida que la muestra aumenta, hasta llegar a $n = 500$ en que dichos valores son casi idénticos.
4. *Estimaciones de los errores estándar de los coeficientes de regresión:* En la Tabla 11 y la Tabla 12 se presentan las estimaciones de los errores estándar de los coeficientes de regresión, y en la Tabla 13 las estimaciones Monte Carlo de los verdaderos errores estándar que sirven de referencia para valorar las estimaciones realizadas por los diferentes modelos. Se puede observar, en la línea de lo que señala Krzanowski (1998), una infraestimación general de los errores estándar por parte del MRP. Por otro lado, las estimaciones con MR Negbin II funcionan correctamente para muestras moderadamente grandes ($n = 100$ y $n = 500$) mientras que con MRQP las estimaciones son

correctas para $n = 500$ mientras que con $n = 100$ pueden considerarse aceptables. En cualquier caso, las estimaciones procedentes de MRPQ producen errores estándar superiores al MRP, de forma que se muestra menos tendente a la infraestimación.

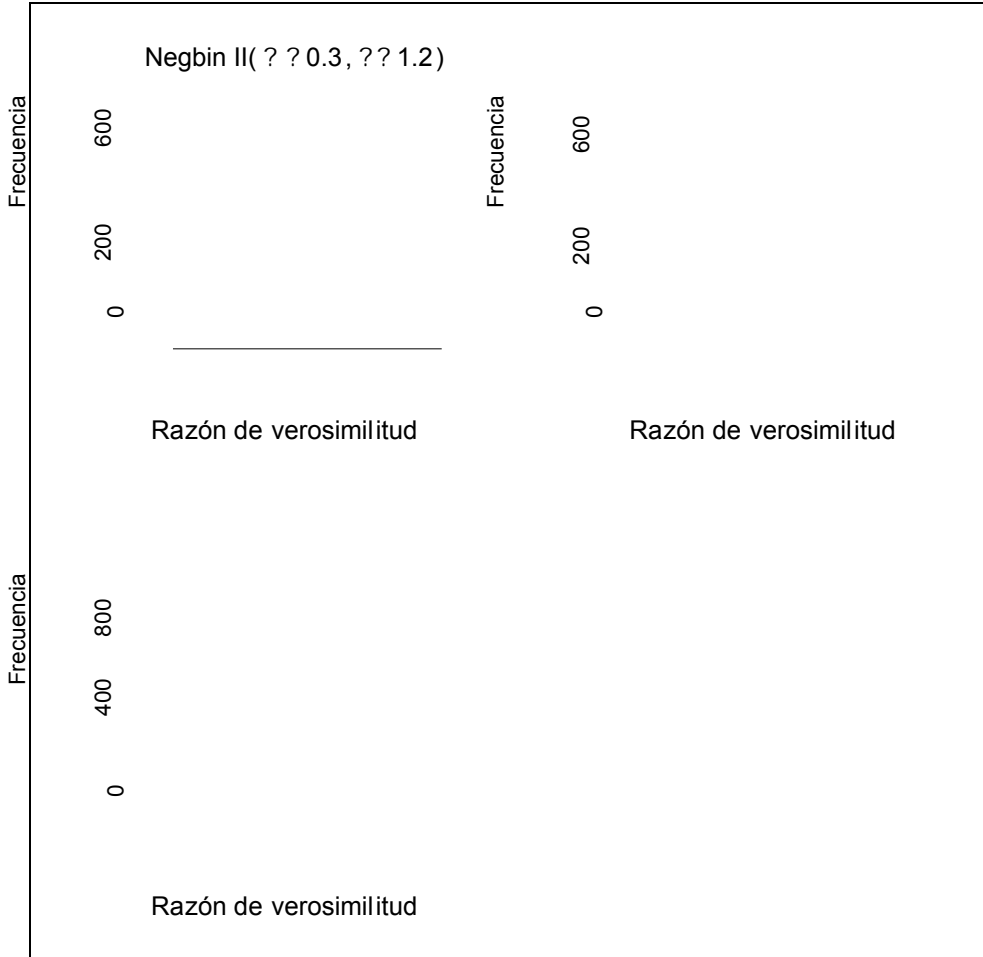


Ilustración 3. Distribuciones muestrales Monte Carlo de la prueba LR para el diagnóstico de la sobredispersión: 5,000 muestras Negbin($\mu=0.3$) de tamaño $n = 500$

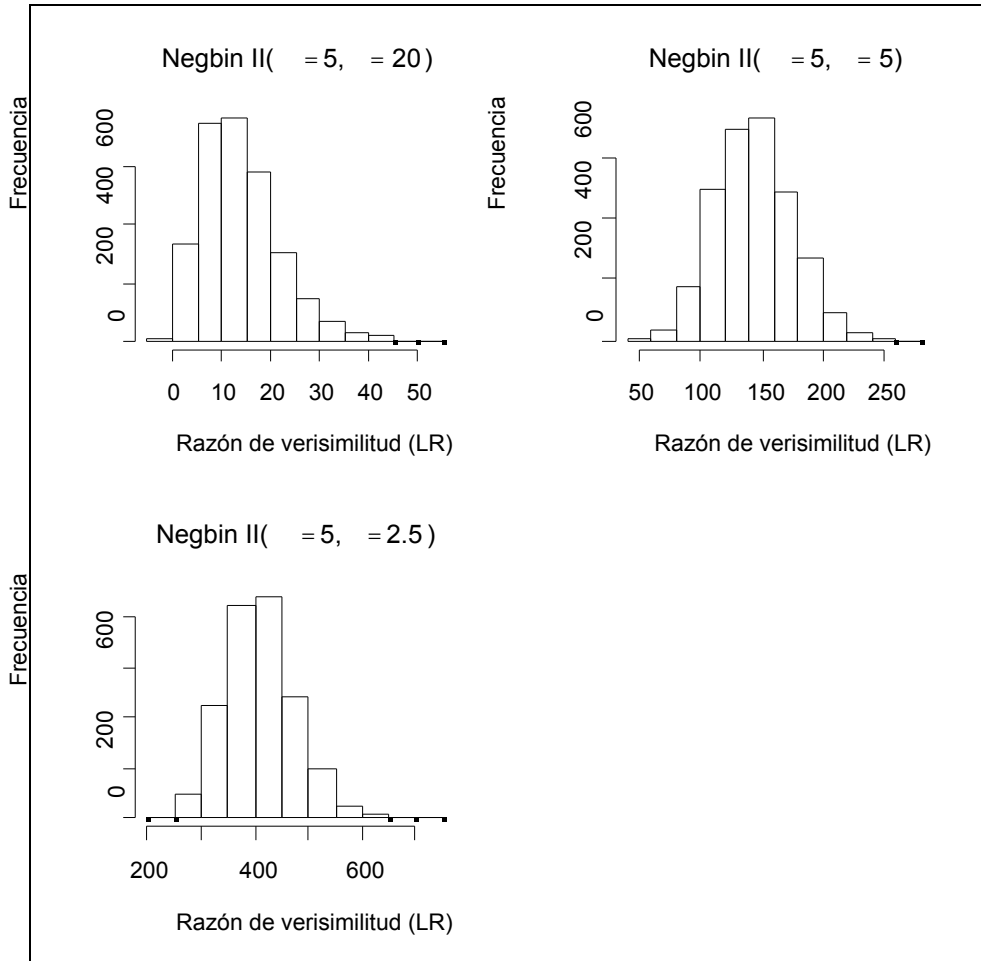


Ilustración 4. Distribuciones muestrales Monte Carlo de la prueba LR para el diagnóstico de la sobredispersión: 5,000 muestras Negbin($\mu = 5$) de tamaño $n = 500$

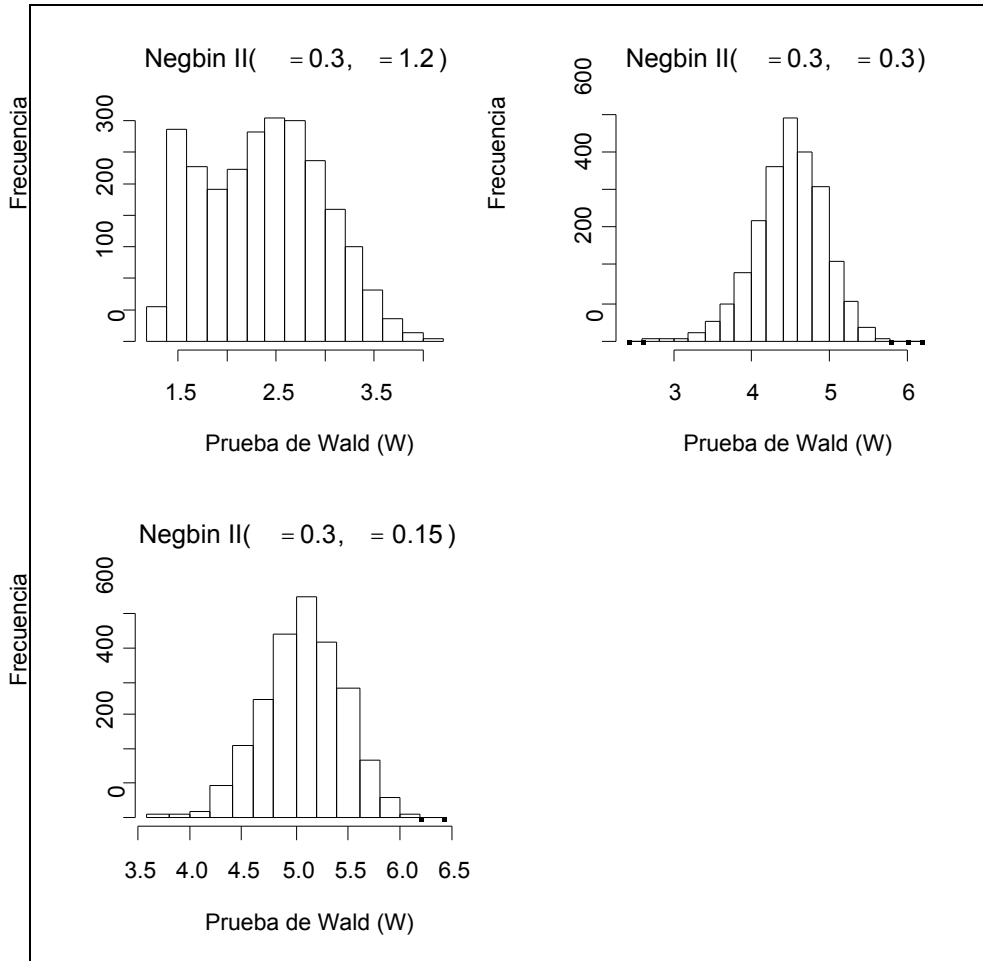


Ilustración 5. Distribuciones muestrales Monte Carlo de la prueba de Wald para el diagnóstico de la sobredispersión: 5,000 muestras Negbin($p = 0.3$) de tamaño $n = 500$

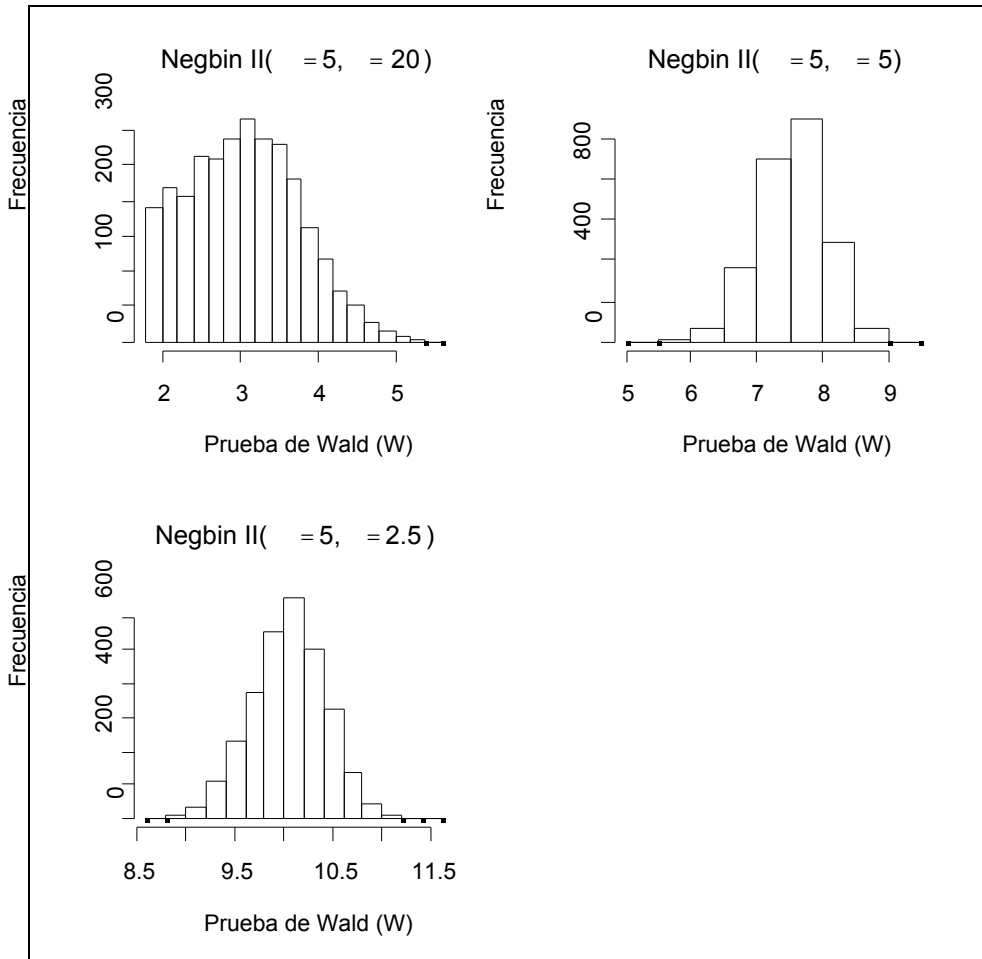


Ilustración 6. Distribuciones muestrales Monte Carlo de la prueba de Wald para el diagnóstico de la sobredispersión: 5,000 muestras Negbin($\mu = 5$) de tamaño $n = 500$

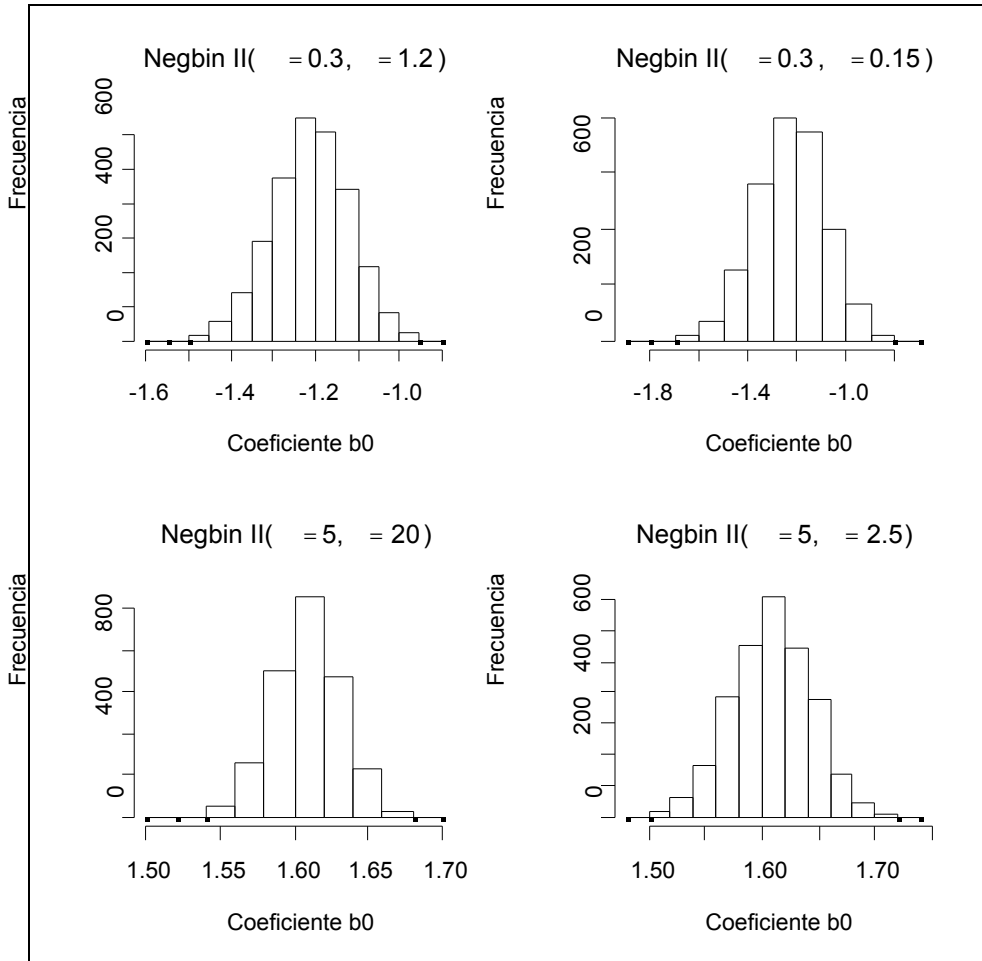


Ilustración 7. Distribuciones muestrales Monte Carlo del coeficiente de regresión MRP b_0 : 5,000 muestras Negbin($\mu = 0.3$ y $\sigma = 5$) de tamaño $n = 500$

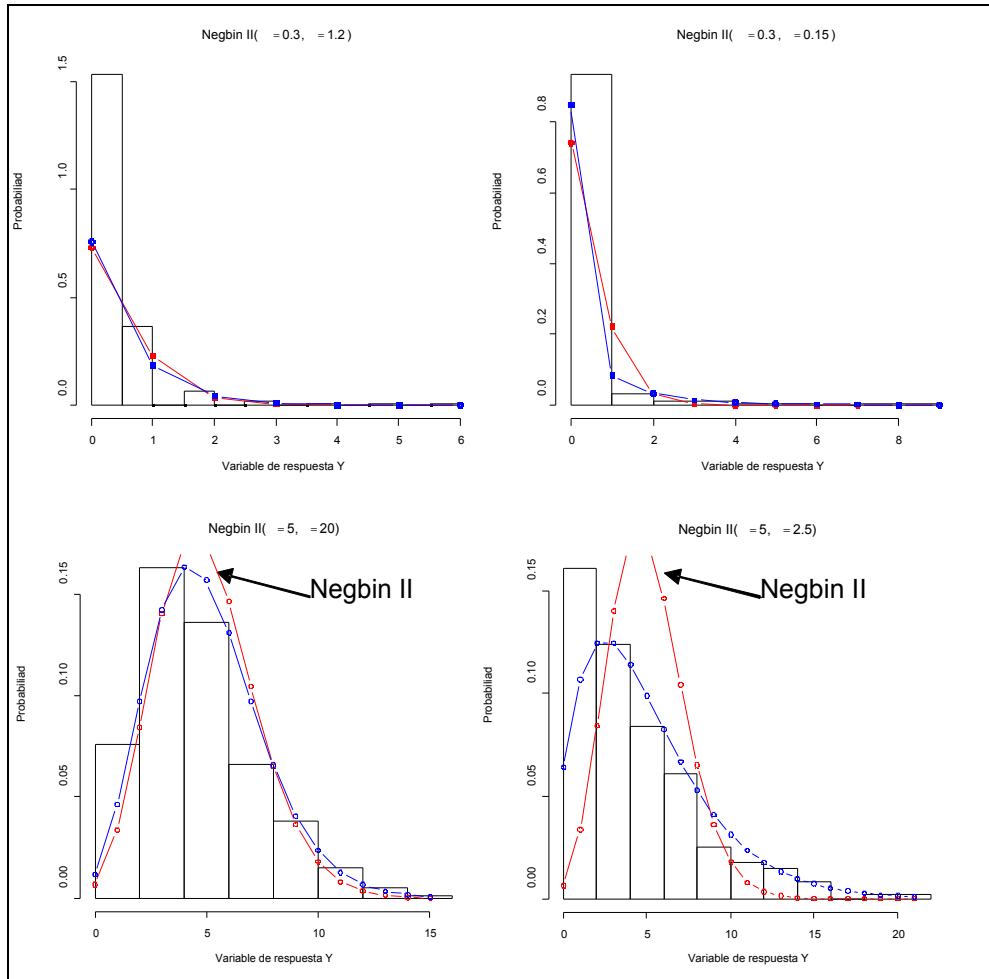


Ilustración 8. Histogramas de sendas muestras de tamaño $n = 500$ generadas bajo cuatro modelos Negbin II con las distribuciones teóricas Negbin y Poisson solapadas.

Tabla 6. Muestras Negbin II:
potencia estadística de los tests diagnósticos de sobredispersión.

N	λ	τ	LR			Wald			LMR Negbin I		
			Pot.	P _{2.5}	P _{97.5}	Pot.	P _{2.5}	P _{97.5}	Pot.	P _{2.5}	P _{97.5}
20	0.3	0.15	0.410	-0.01	25.29	0.000	0.44	1.47	0.033	-0.19	2.19
20	0.3	0.3	0.273	-0.02	15.83	0.000	0.42	1.37	0.025	-0.50	2.10
20	0.3	1.2	0.078	-0.08	5.17	0.000	0.42	1.07	0.010	-0.99	1.71
20	1	0.5	0.742	2.54	34.94	0.000	1.02	1.92	0.157	1.13	2.83
20	1	1	0.453	0.35	17.61	0.000	0.60	1.75	0.083	0.59	2.49
20	1	4	0.098	-0.09	5.76	0.000	0.47	1.36	0.047	-0.72	1.84
20	5	2.5	0.864	5.25	42.82	0.001	1.46	2.46	0.380	1.45	3.29
20	5	5	0.567	0.95	20.63	0.000	0.79	2.15	0.148	0.86	2.75
20	5	20	0.110	-0.08	5.90	0.000	0.50	1.54	0.070	-0.55	1.88
50	0.3	0.15	0.839	5.02	56.92	0.000	1.14	2.11	0.109	1.14	2.44
50	0.3	0.3	0.668	1.68	30.80	0.000	0.88	1.98	0.099	0.98	2.40
50	0.3	1.2	0.189	-0.08	8.65	0.000	0.57	1.53	0.030	-0.21	1.90
50	1	0.5	0.980	14.91	70.55	0.133	2.06	2.93	0.527	1.69	3.44
50	1	1	0.850	4.46	34.37	0.012	1.48	2.62	0.383	1.43	3.07
50	1	4	0.214	-0.05	9.77	0.000	0.68	1.90	0.058	-0.00	2.21
50	5	2.5	0.999	24.44	81.50	0.801	2.79	3.70	0.908	2.42	4.38
50	5	5	0.918	6.28	35.93	0.168	1.82	3.10	0.635	1.80	3.58
50	5	20	0.244	-0.01	8.93	0.000	0.71	2.06	0.076	0.12	2.35
100	0.3	0.15	0.988	33.18	167.76	0.085	1.91	2.89	0.244	1.41	2.60
100	0.3	0.3	0.918	7.16	47.64	0.013	1.58	2.62	0.334	1.44	2.79
100	0.3	1.2	0.346	0.12	13.10	0.000	0.76	1.97	0.076	0.35	2.32
100	1	0.5	1.000	87.26	235.09	0.993	3.46	4.25	0.801	2.07	3.76
100	1	1	0.982	13.51	54.50	0.539	2.40	3.51	0.789	2.06	3.73
100	1	4	0.409	0.31	12.91	0.001	0.92	2.39	0.145	0.56	2.58
100	5	2.5	1.000	157.45	314.64	1.000	5.13	5.79	0.994	3.39	5.82
100	5	5	0.997	18.02	58.89	0.841	2.91	4.18	0.965	2.73	4.52
100	5	20	0.416	0.44	12.17	0.012	0.97	2.55	0.173	0.64	2.69
500	0.3	0.15	1.000	146.24	306.77	1.000	4.833	5.78	0.973	2.810	4.460
500	0.3	0.3	1.000	65.23	154.16	0.999	4.833	5.78	0.973	2.810	4.460
500	0.3	1.2	0.888	5.28	30.65	0.312	4.231	5.33	0.987	2.968	4.567
500	1	0.5	1.000	258.93	427.78	1.000	7.541	8.43	0.999	4.412	6.581
500	1	1	1.000	99.96	188.53	1.000	6.047	7.22	1.000	4.484	6.396
500	1	4	0.943	6.98	30.97	0.559	2.287	4.14	0.847	2.239	4.087
500	5	2.5	1.000	364.08	537.74	1.000	9.794	10.73	1.000	7.304	9.765
500	5	5	1.000	121.09	209.76	1.000	7.185	8.48	1.000	6.324	8.301
500	5	20	0.960	7.83	32.65	0.660	2.495	4.49	0.895	2.464	4.514

Tabla 7. (continuación...) Muestras Negbin II:
potencia estadística de los tests diagnósticos de sobredispersión.

N	λ	τ	LMR Negbin II			Discrepancia			χ^2		
			Pot.	P _{2.5}	P _{97.5}	Pot.	P _{2.5}	P _{97.5}	Pot.	P _{2.5}	P _{97.5}
20	0.3	0.15	0.152	-0.01	3.34	0.200	11.08	51.79	0.492	17.24	99.67
20	0.3	0.3	0.097	-0.45	2.98	0.121	11.94	40.89	0.353	17.19	71.07
20	0.3	1.2	0.036	-1.09	1.99	0.015	12.58	27.48	0.125	15.50	39.31
20	1	0.5	0.227	1.13	3.25	0.762	29.35	73.09	0.796	31.13	103.97
20	1	1	0.120	0.61	2.80	0.568	24.80	53.68	0.548	23.27	67.36
20	1	4	0.050	-0.70	1.92	0.205	19.78	37.81	0.165	16.76	40.62
20	5	2.5	0.394	1.45	3.40	0.932	39.71	94.48	0.908	37.47	100.74
20	5	5	0.161	0.85	2.82	0.728	28.20	64.22	0.674	26.65	66.97
20	5	20	0.071	-0.54	1.91	0.239	18.13	40.87	0.195	17.27	40.39
50	0.3	0.15	0.210	1.18	2.98	0.336	41.93	109.74	0.862	77.58	270.35
50	0.3	0.3	0.161	1.01	2.83	0.177	40.00	86.43	0.713	63.03	166.55
50	0.3	1.2	0.050	-0.23	2.07	0.014	37.96	62.12	0.268	48.22	93.68
50	1	0.5	0.555	1.72	3.54	0.978	89.15	160.45	0.985	102.24	245.48
50	1	1	0.404	1.44	3.20	0.913	74.06	120.75	0.884	74.29	157.99
50	1	4	0.063	0.00	2.23	0.434	57.22	86.09	0.296	49.97	91.23
50	5	2.5	0.908	2.44	4.41	1.000	120.74	200.93	0.999	117.33	222.84
50	5	5	0.638	1.79	3.63	0.965	82.96	139.09	0.942	79.15	141.24
50	5	20	0.078	0.12	2.35	0.398	53.47	89.32	0.317	51.08	86.87
100	0.3	0.15	0.300	1.44	2.89	0.599	103.15	266.47	0.991	260.75	952.73
100	0.3	0.3	0.371	1.46	2.95	0.227	90.44	152.91	0.933	147.28	312.84
100	0.3	1.2	0.086	0.37	2.39	0.010	81.16	115.22	0.411	104.40	176.10
100	1	0.5	0.811	2.10	3.82	1.000	241.04	414.29	1.000	353.65	812.98
100	1	1	0.795	2.07	3.75	0.992	158.65	223.46	0.985	164.57	277.80
100	1	4	0.149	0.56	2.59	0.723	120.98	160.47	0.473	108.15	166.08
100	5	2.5	0.994	3.40	5.84	1.000	399.96	589.09	1.000	415.07	708.38
100	5	5	0.967	2.72	4.55	0.100	178.35	257.47	0.998	173.22	262.32
100	5	20	0.176	0.64	2.71	0.598	114.89	164.33	0.484	109.80	158.99
500	0.3	0.15	0.974	2.834	4.51	0.846	576.6	794.0	1.000	1253.7	2068.0
500	0.3	0.3	0.974	2.834	4.51	0.846	576.6	794.0	1.000	1253.7	2068.0
500	0.3	1.2	0.989	2.987	4.59	0.459	513.6	657.6	1.000	888.5	1322.1
500	1	0.5	0.999	4.434	6.60	1.000	1061.1	1285.5	1.000	1350.2	1866.5
500	1	1	1.000	4.489	6.40	1.000	854.5	1000.7	1.000	929.4	1204.9
500	1	4	0.850	2.232	4.07	1.000	641.4	727.6	0.953	590.2	712.8
500	5	2.5	1.000	7.308	9.78	1.000	1400.7	1649.8	1.000	1410.2	1737.0
500	5	5	1.000	6.329	8.31	1.000	964.5	1136.4	1.000	944.3	1137.3
500	5	20	0.894	2.466	4.49	0.994	618.9	731.3	0.967	593.9	708.2

Tabla 8. Muestras Negbin II:
estimaciones del parámetro de dispersión.

N	λ	τ	Negbin II			Quasi-Poisson		
			α	P _{2.5}	P _{97.5}	ϕ	P _{2.5}	P _{97.5}
20	0.3	0.15	41.82	0.2005	67.9925	1.863	0.9547	5.534
20	0.3	0.3	37.94	0.3306	44.8765	1.529	0.9532	3.831
20	0.3	1.2	8.755	1.1281	29.9268	1.135	0.8568	2.177
20	1	0.5	2.170	0.4340	17.6636	2.580	1.7283	5.755
20	1	1	6.387	0.8545	44.3369	1.825	1.2891	3.740
20	1	4	21.81	2.9333	78.4171	1.219	0.9269	2.254
20	5	2.5	6.436	2.2126	31.9874	2.880	2.0811	5.578
20	5	5	22.61	4.2096	184.0520	1.985	1.4807	3.718
20	5	20	103.5	13.3711	383.3493	1.262	0.9578	2.243
50	0.3	0.15	0.8883	0.1282	4.4340	2.428	1.6122	5.629
50	0.3	0.3	1.075	0.2490	7.7168	1.777	1.3113	3.455
50	0.3	1.2	4.491	0.8209	16.3733	1.217	1.0043	1.944
50	1	0.5	0.6743	0.4277	1.7679	2.798	2.1288	5.093
50	1	1	1.898	0.8160	11.0765	1.923	1.5416	3.290
50	1	4	12.330	2.7257	46.4719	1.234	1.0369	1.894
50	5	2.5	2.996	2.1831	6.0811	2.960	2.4429	4.627
50	5	5	7.928	4.2920	26.6122	1.970	1.6485	2.941
50	5	20	58.150	13.7551	226.3825	1.252	1.0643	1.809
100	0.3	0.15	1.99e+07	0.0639	0.2745	4.081	2.6405	9.663
100	0.3	0.3	0.5004	0.2496	1.9420	1.882	1.5003	3.169
100	0.3	1.2	3.158	0.8326	12.7352	1.233	1.0649	1.794
100	1	0.5	0.2763	0.2218	0.4583	4.657	3.6011	8.235
100	1	1	1.265	0.8529	2.9072	1.958	1.6729	2.833
100	1	4	8.696	2.8206	34.1158	1.240	1.0984	1.687
100	5	2.5	1.321	1.1374	1.9447	4.920	4.2231	7.194
100	5	5	5.841	4.3742	11.3591	1.991	1.7674	2.676
100	5	20	39.94	14.9047	153.4430	1.246	1.1204	1.622
500	0.3	0.15	0.1579	0.1353	0.2254	2.879	2.505	4.099
500	0.3	0.3	0.3162	0.2646	0.4891	1.978	1.781	2.640
500	0.3	1.2	1.5540	0.9473	4.4668	1.247	1.166	1.485
500	1	0.5	0.5133	0.4642	0.6625	2.961	2.710	3.743
500	1	1	1.0276	0.9171	1.4042	1.995	1.865	2.419
500	1	4	4.7799	3.3406	11.0337	1.239	1.179	1.421
500	5	2.5	2.5366	2.3591	3.0871	2.998	2.830	3.485
500	5	5	5.1480	4.6398	6.7796	1.996	1.896	2.283
500	5	20	23.1751	16.7340	51.0008	1.250	1.193	1.422

Tabla 9. Muestras Negbin II:
estimaciones del coeficiente b_0 del modelo de regresión para recuentos.

N	λ	τ	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	-4.448	-2.423	-0.333	-2.918	-2.365	-0.316	-2.918	-2.365	-0.316
20	0.3	0.3	-2.831	-2.025	-0.435	-2.122	-1.971	-0.431	-2.122	-1.971	-0.431
20	0.3	1.2	-1.995	-1.755	-0.563	-1.681	-1.757	-0.558	-1.681	-1.757	-0.558
20	1	0.5	-0.167	-0.432	0.570	-0.163	-0.431	0.571	-0.163	-0.431	0.571
20	1	1	-0.108	-0.317	0.489	-0.106	-0.316	0.489	-0.106	-0.316	0.489
20	1	4	-0.065	-0.227	0.399	-0.064	-0.227	0.399	-0.064	-0.227	0.399
20	5	2.5	1.580	1.464	1.904	1.580	1.463	1.903	1.580	1.463	1.903
20	5	5	1.589	1.500	1.847	1.589	1.499	1.847	1.589	1.499	1.847
20	5	20	1.596	1.520	1.8077	1.596	1.519	1.808	1.596	1.519	1.808
50	0.3	0.15	-1.442	-1.739	-0.581	-1.433	-1.730	-0.580	-1.433	-1.730	-0.580
50	0.3	0.3	-1.371	-1.611	-0.654	-1.367	-1.610	-0.654	-1.367	-1.610	-0.654
50	0.3	1.2	-1.297	-1.487	-0.742	-1.296	-1.485	-0.742	-1.296	-1.485	-0.742
50	1	0.5	-0.063	-0.228	0.392	-0.062	-0.228	0.392	-0.062	-0.228	0.392
50	1	1	-0.047	-0.176	0.344	-0.047	-0.176	0.345	-0.047	-0.176	0.345
50	1	4	-0.030	-0.137	0.269	-0.029	-0.137	0.270	-0.029	-0.137	0.270
50	5	2.5	1.598	1.521	1.820	1.598	1.522	1.819	1.598	1.522	1.819
50	5	5	1.599	1.538	1.771	1.599	1.538	1.771	1.599	1.538	1.771
50	5	20	1.605	1.559	1.744	1.605	1.559	1.744	1.605	1.559	1.744
100	0.3	0.15	-1.376	-1.662	-0.602	-1.366	-1.634	-0.602	-1.366	-1.634	-0.602
100	0.3	0.3	-1.276	-1.454	-0.798	-1.275	-1.500	-0.799	-1.275	-1.450	-0.799
100	0.3	1.2	-1.249	-1.385	-0.865	-1.249	-1.385	-0.865	-1.249	-1.385	-0.865
100	1	0.5	-0.045	-0.185	0.363	-0.044	-0.186	0.3647	-0.044	-0.186	0.365
100	1	1	-0.021	-0.120	0.257	-0.021	-0.120	0.257	-0.021	-0.120	0.257
100	1	4	-0.010	-0.082	0.196	-0.010	-0.082	0.196	-0.010	-0.082	0.196
100	5	2.5	1.600	1.532	1.791	1.600	1.532	1.793	1.600	1.532	1.793
100	5	5	1.607	1.565	1.730	1.607	1.565	1.730	1.607	1.565	1.730
100	5	20	1.607	1.573	1.704	1.607	1.573	1.704	1.606	1.573	1.704
500	0.3	0.15	-1.228	-1.321	-0.965	-1.228	-1.321	-0.963	-1.228	-1.321	-0.963
500	0.3	0.3	-1.216	-1.295	-0.989	-1.216	-1.295	-0.989	-1.216	-1.295	-0.989
500	0.3	1.2	-1.212	-1.273	-1.038	-1.213	-1.273	-1.038	-1.213	-1.274	-1.039
500	1	0.5	-0.009	-0.059	0.142	-0.009	-0.059	0.142	-0.008	-0.059	0.142
500	1	1	-0.004	-0.047	0.120	-0.004	-0.048	0.119	-0.004	-0.047	0.119
500	1	4	-0.002	-0.036	0.095	-0.003	-0.036	0.095	-0.003	-0.036	0.095
500	5	2.5	1.609	1.585	1.674	1.609	1.585	1.674	1.609	1.585	1.674
500	5	5	1.608	1.590	1.664	1.608	1.590	1.664	1.608	1.590	1.663
500	5	20	1.609	1.594	1.651	1.609	1.594	1.651	1.609	1.594	1.651

Tabla 10. Muestras Negbin II:
estimaciones del coeficiente b_1 del modelo de regresión para recuentos.

N	λ	τ	MRNegbin II			MRP			MRQP		
			b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	-0.066	-0.297	1.423	-0.024	-0.248	1.281	-0.024	-0.248	1.281
20	0.3	0.3	-0.137	-0.249	1.075	-0.052	-0.225	0.894	-0.052	-0.225	0.894
20	0.3	1.2	-0.028	-0.191	0.728	-0.016	-0.183	0.660	-0.016	-0.183	0.660
20	1	0.5	0.005	-0.153	0.515	0.003	-0.145	0.478	0.003	-0.145	0.478
20	1	1	-0.004	-0.123	0.376	-0.003	-0.120	0.367	-0.003	-0.129	0.367
20	1	4	0.004	-0.094	0.306	0.003	-0.093	0.305	0.003	-0.093	0.305
20	5	2.5	0.002	-0.066	0.192	0.002	-0.065	0.187	0.002	-0.065	0.187
20	5	5	0.000	-0.056	0.160	0.001	-0.055	0.159	0.000	-0.055	0.159
20	5	20	-0.001	-0.043	0.124	-0.001	-0.043	0.123	-0.001	-0.043	0.123
50	0.3	0.15	0.005	-0.186	0.650	0.005	-0.170	0.616	0.005	-0.170	0.616
50	0.3	0.3	0.005	-0.146	0.513	0.006	-0.140	0.493	0.006	-0.140	0.493
50	0.3	1.2	0.000	-0.114	0.371	0.001	-0.113	0.366	0.001	-0.1127	0.366
50	1	0.5	-0.002	-0.096	0.298	-0.002	-0.094	0.289	-0.002	-0.094	0.289
50	1	1	0.001	-0.076	0.240	0.001	-0.075	0.239	0.001	-0.075	0.238
50	1	4	-0.001	-0.065	0.190	-0.001	-0.065	0.185	-0.001	-0.065	0.185
50	5	2.5	0.001	-0.041	0.126	0.001	-0.041	0.125	0.001	-0.041	0.125
50	5	5	-0.001	-0.034	0.100	-0.001	-0.034	0.100	-0.001	-0.034	0.100
50	5	20	-0.001	-0.027	0.078	-0.001	-0.027	0.077	-0.001	-0.027	0.077
100	0.3	0.15	0.019	-0.291	1.059	0.038	-0.235	0.944	0.038	-0.236	0.944
100	0.3	0.3	0.009	-0.171	0.609	0.012	-0.163	0.590	0.012	-0.163	0.590
100	0.3	1.2	0.004	-0.143	0.431	0.005	-0.144	0.427	0.005	-0.144	0.427
100	1	0.5	0.013	-0.150	0.509	0.017	-0.139	0.489	0.017	-0.139	0.489
100	1	1	0.002	-0.099	0.304	0.002	-0.097	0.301	0.002	-0.097	0.301
100	1	4	0.000	-0.076	0.232	0.000	-0.076	0.231	0.000	-0.076	0.231
100	5	2.5	0.002	-0.071	0.220	0.002	-0.071	0.210	0.002	-0.071	0.210
100	5	5	0.002	-0.044	0.133	0.002	-0.044	0.132	0.002	-0.044	0.133
100	5	20	0.001	-0.034	0.105	0.001	-0.035	0.106	0.001	-0.035	0.106
500	0.3	0.15	-0.001	-0.056	0.162	-0.001	-0.056	0.162	-0.001	-0.056	0.162
500	0.3	0.3	0.002	-0.044	0.142	0.002	-0.044	0.138	0.002	-0.044	0.138
500	0.3	1.2	0.000	-0.034	0.104	0.001	-0.034	0.104	0.001	-0.034	0.104
500	1	0.5	-0.000	-0.031	0.088	-0.001	-0.031	0.088	-0.001	-0.031	0.088
500	1	1	0.000	-0.022	0.071	0.001	-0.023	0.070	0.001	-0.023	0.070
500	1	4	0.000	-0.019	0.057	0.000	-0.019	0.057	0.001	-0.019	0.057
500	5	2.5	-0.000	-0.014	0.037	-0.001	-0.014	0.037	-0.001	-0.014	0.037
500	5	5	-0.000	-0.011	0.031	-0.000	-0.011	0.031	-0.000	-0.011	0.031
500	5	20	0.000	-0.008	0.025	0.000	-0.008	0.025	0.000	-0.008	0.025

Tabla 11. Muestras Negbin II: estimaciones de los errores estándar del coeficiente b_0 del modelo de regresión para recuentos.

n	λ	τ	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	193.113	0.585	470.326	4.219	0.405	63.652	0.769	0.546	1.833
20	0.3	0.3	89.040	0.493	6.105	1.972	0.388	6.046	0.666	0.478	1.660
20	0.3	1.2	41.221	0.419	1.688	1.133	0.380	1.603	0.533	0.401	1.123
20	1	0.5	0.392	0.323	0.643	0.260	0.213	0.437	0.395	0.325	0.657
20	1	1	0.317	0.268	0.484	0.246	0.210	0.383	0.321	0.269	0.488
20	1	4	0.261	0.228	0.372	0.237	0.211	0.320	0.257	0.218	0.379
20	5	2.5	0.164	0.143	0.225	0.103	0.096	0.123	0.170	0.148	0.237
20	5	5	0.135	0.118	0.185	0.102	0.097	0.118	0.140	0.123	0.195
20	5	20	0.112	0.102	0.147	0.101	0.097	0.113	0.112	0.098	0.152
50	0.3	0.15	0.479	0.389	0.824	0.319	0.244	0.594	0.464	0.376	0.807
50	0.3	0.3	0.390	0.326	0.657	0.300	0.244	0.522	0.384	0.322	0.633
50	0.3	1.2	0.310	0.269	0.461	0.281	0.244	0.414	0.306	0.263	0.455
50	1	0.5	0.244	0.220	0.326	0.149	0.135	0.196	0.244	0.217	0.334
50	1	1	0.200	0.179	0.262	0.147	0.136	0.186	0.201	0.179	0.268
50	1	4	0.161	0.147	0.201	0.145	0.136	0.174	0.160	0.145	0.203
50	5	2.5	0.107	0.099	0.132	0.064	0.061	0.071	0.109	0.100	0.136
50	5	5	0.087	0.081	0.107	0.064	0.062	0.070	0.089	0.082	0.109
50	5	20	0.070	0.066	0.085	0.064	0.062	0.068	0.071	0.065	0.086
100	0.3	0.15	0.418	0.358	0.615	0.209	0.171	0.340	0.395	0.336	0.601
100	0.3	0.3	0.264	0.237	0.356	0.194	0.173	0.264	0.261	0.233	0.358
100	0.3	1.2	0.211	0.192	0.273	0.190	0.175	0.240	0.210	0.190	0.274
100	1	0.5	0.223	0.206	0.277	0.104	0.095	0.132	0.220	0.198	0.285
100	1	1	0.141	0.132	0.170	0.102	0.096	0.118	0.142	0.131	0.173
100	1	4	0.113	0.106	0.133	0.101	0.097	0.113	0.112	0.104	0.134
100	5	2.5	0.100	0.094	0.115	0.046	0.044	0.050	0.100	0.093	0.119
100	5	5	0.063	0.059	0.072	0.045	0.044	0.047	0.063	0.059	0.073
100	5	20	0.050	0.047	0.057	0.045	0.044	0.047	0.050	0.047	0.057
500	0.3	0.15	0.141	0.134	0.162	0.083	0.078	0.095	0.139	0.131	0.166
500	0.3	0.3	0.115	0.110	0.130	0.082	0.079	0.092	0.115	0.109	0.133
500	0.3	1.2	0.091	0.088	0.102	0.082	0.079	0.090	0.091	0.088	0.103
500	1	0.5	0.077	0.074	0.084	0.045	0.043	0.048	0.077	0.074	0.086
500	1	1	0.063	0.061	0.068	0.044	0.043	0.047	0.063	0.061	0.069
500	1	4	0.049	0.048	0.053	0.044	0.043	0.047	0.049	0.048	0.053
500	5	2.5	0.034	0.033	0.036	0.020	0.019	0.020	0.034	0.033	0.037
500	5	5	0.028	0.027	0.030	0.020	0.019	0.020	0.028	0.027	0.030
500	5	20	0.022	0.021	0.023	0.020	0.019	0.020	0.022	0.021	0.023

Tabla 12. Muestras Negbin II:
estimaciones de los errores estándar del coeficiente b_1
del modelo de regresión para recuentos.

n	λ	τ	MR Negbin II			MRP			MRQP		
			b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	68.274	0.319	258.288	2.013	0.211	34.955	0.376	0.288	0.735
20	0.3	0.3	31.062	0.269	2.245	0.925	0.207	2.223	0.335	0.256	0.646
20	0.3	1.2	14.388	0.226	0.777	0.545	0.207	0.679	0.275	0.215	0.565
20	1	0.5	0.214	0.176	0.348	0.137	0.114	0.226	0.209	0.174	0.335
20	1	1	0.173	0.145	0.263	0.132	0.114	0.195	0.172	0.146	0.255
20	1	4	0.141	0.124	0.202	0.128	0.114	0.170	0.139	0.119	0.203
20	5	2.5	0.090	0.078	0.123	0.056	0.052	0.067	0.092	0.081	0.130
20	5	5	0.074	0.064	0.101	0.055	0.052	0.064	0.077	0.067	0.106
20	5	20	0.061	0.055	0.080	0.055	0.053	0.061	0.061	0.053	0.083
50	0.3	0.15	0.268	0.219	0.458	0.171	0.135	0.304	0.250	0.207	0.402
50	0.3	0.3	0.218	0.183	0.357	0.164	0.136	0.269	0.210	0.179	0.329
50	0.3	1.2	0.173	0.150	0.253	0.155	0.137	0.222	0.169	0.148	0.246
50	1	0.5	0.138	0.123	0.184	0.083	0.075	0.108	0.136	0.121	0.185
50	1	1	0.112	0.101	0.148	0.082	0.076	0.103	0.112	0.100	0.149
50	1	4	0.090	0.083	0.113	0.081	0.076	0.097	0.089	0.081	0.114
50	5	2.5	0.060	0.056	0.074	0.036	0.034	0.040	0.061	0.056	0.077
50	5	5	0.049	0.045	0.060	0.036	0.034	0.039	0.050	0.046	0.061
50	5	20	0.039	0.037	0.047	0.035	0.035	0.038	0.039	0.036	0.048
100	0.3	0.15	0.436	0.372	0.650	0.203	0.168	0.324	0.385	0.329	0.568
100	0.3	0.3	0.271	0.242	0.368	0.194	0.174	0.263	0.261	0.234	0.355
100	0.3	1.2	0.215	0.195	0.279	0.192	0.176	0.244	0.212	0.192	0.274
100	1	0.5	0.231	0.212	0.286	0.105	0.096	0.135	0.222	0.200	0.291
100	1	1	0.145	0.135	0.175	0.104	0.098	0.121	0.144	0.133	0.176
100	1	4	0.115	0.108	0.137	0.103	0.099	0.116	0.114	0.107	0.137
100	5	2.5	0.102	0.097	0.118	0.046	0.044	0.051	0.102	0.095	0.122
100	5	5	0.064	0.061	0.074	0.046	0.045	0.049	0.065	0.061	0.076
100	5	20	0.051	0.048	0.058	0.046	0.045	0.048	0.051	0.048	0.059
500	0.3	0.15	0.081	0.077	0.093	0.047	0.045	0.054	0.080	0.075	0.095
500	0.3	0.3	0.066	0.063	0.075	0.047	0.045	0.053	0.066	0.063	0.076
500	0.3	1.2	0.052	0.050	0.058	0.047	0.045	0.051	0.052	0.050	0.059
500	1	0.5	0.044	0.043	0.048	0.025	0.025	0.027	0.044	0.042	0.049
500	1	1	0.036	0.035	0.039	0.025	0.025	0.027	0.036	0.035	0.039
500	1	4	0.028	0.027	0.031	0.025	0.025	0.027	0.028	0.027	0.030
500	5	2.5	0.019	0.019	0.021	0.011	0.011	0.011	0.019	0.019	0.021
500	5	5	0.016	0.015	0.017	0.011	0.011	0.011	0.016	0.015	0.017
500	5	20	0.012	0.012	0.013	0.011	0.011	0.011	0.012	0.012	0.013

Tabla 13. Muestras Negbin II: estimaciones Monte Carlo de los errores estándar de los coeficientes de regresión.

n	λ	τ	b_0			b_1		
			MR Negbin II	MRP	MRQP	MR Negbin II	MRP	MRQP
20	0.3	0.15	17.476	6.467	6.467	5.924	2.206	2.206
20	0.3	0.3	12.072	4.480	4.480	4.076	1.538	1.538
20	0.3	1.2	8.095	3.002	3.002	2.724	1.030	1.030
20	1	0.5	0.452	0.450	0.450	0.255	0.237	0.237
20	1	1	0.360	0.359	0.359	0.195	0.188	0.188
20	1	4	0.274	0.274	0.274	0.147	0.145	0.145
20	5	2.5	0.179	0.179	0.179	0.098	0.096	0.096
20	5	5	0.143	0.143	0.143	0.079	0.079	0.079
20	5	20	0.113	0.113	0.113	0.063	0.062	0.062
50	0.3	0.15	0.571	0.562	0.562	0.325	0.298	0.298
50	0.3	0.3	0.470	0.464	0.464	0.250	0.237	0.237
50	0.3	1.2	0.333	0.332	0.332	0.182	0.180	0.180
50	1	0.5	0.254	0.255	0.255	0.147	0.142	0.142
50	1	1	0.209	0.209	0.209	0.119	0.117	0.117
50	1	4	0.167	0.167	0.167	0.094	0.093	0.093
50	5	2.5	0.111	0.111	0.111	0.063	0.062	0.062
50	5	5	0.091	0.091	0.091	0.051	0.051	0.051
50	5	20	0.071	0.071	0.071	0.039	0.039	0.039
100	0.3	0.15	0.523	0.450	0.450	0.594	0.410	0.410
100	0.3	0.3	0.280	0.279	0.279	0.287	0.273	0.273
100	0.3	1.2	0.220	0.219	0.219	0.219	0.216	0.216
100	1	0.5	0.230	0.229	0.229	0.247	0.232	0.232
100	1	1	0.147	0.147	0.147	0.150	0.148	0.148
100	1	4	0.111	0.111	0.111	0.116	0.116	0.116
100	5	2.5	0.100	0.100	0.100	0.107	0.105	0.105
100	5	5	0.064	0.064	0.064	0.066	0.065	0.065
100	5	20	0.051	0.051	0.051	0.051	0.051	0.051
500	0.3	0.15	0.142	0.142	0.142	0.083	0.082	0.082
500	0.3	0.3	0.118	0.118	0.118	0.069	0.069	0.069
500	0.3	1.2	0.092	0.092	0.092	0.053	0.053	0.053
500	1	0.5	0.077	0.077	0.077	0.045	0.045	0.045
500	1	1	0.064	0.064	0.064	0.035	0.035	0.035
500	1	4	0.050	0.050	0.050	0.028	0.028	0.028
500	5	2.5	0.034	0.034	0.034	0.019	0.019	0.019
500	5	5	0.027	0.027	0.027	0.016	0.016	0.016
500	5	20	0.022	0.022	0.022	0.012	0.012	0.012

5.2.2 Heterogeneidad no observada simulada mediante la mezcla de dos distribuciones de Poisson

Para simular la heterogeneidad se ha recurrido a una mezcla de dos distribuciones de Poisson con diferente parámetro de localización, para reproducir otro de los mecanismos generadores de heterogeneidad no observada más habituales (nótese que entre el segundo y el tercer caso existe la misma diferencia entre λ_1 y λ_2):

λ_1	λ_2
0.2	1
0.2	3
3.2	6

En cada una de las 5,000 muestras generadas, esta combinación de valores se repite para cada uno de los tamaños muestrales utilizados:

$$n = 500, n = 100, n = 50 \text{ y } n = 20$$

1. *Potencia estadística de las pruebas diagnósticas de sobredispersión:* Se observa una cierta mejora en la potencia a medida que aumenta la sobredispersión y el tamaño muestral, sin embargo no es una tendencia constante sino que se muestra algo errática (ver Tabla 14). La excepción es la prueba de Wald en la que hay una evidente falta de potencia para cualquier combinación de n , λ_1 y λ_2 .
2. *Estimaciones de los parámetros de dispersión:* En todos los tamaños muestrales las estimaciones de λ mediante quasi-Poisson son muy estables, mientras que las del parámetro de dispersión k de Negbin II aparecen como más ineficientes (ver Tabla 15).
3. *Estabilidad de las estimaciones de β_0 y β_1 a lo largo de los diferentes modelos de regresión ajustados:* Las estimaciones de los modelos MRP y MRQP son idénticas en todos los valores y para ambos parámetros (β_0 y β_1). En cuanto a Negbin II muestra valores iguales a los otros dos modelos cuando $\lambda_1 = 3.2$ y $\lambda_2 = 6$. Esto parece indicar que lo que influye en las estimaciones no es tanto la diferencia entre valores esperados de ambas distribuciones, sino los propios valores esperados (véase Tabla 16)
4. *Estimaciones de los errores estándar de los coeficientes de regresión:* En la Tabla 17 se presentan las estimaciones de los errores estándar de los coeficientes de regresión; en la Tabla 18 se presentan las estimaciones Monte Carlo de los verdaderos errores estándar que sirven de referencia para valorar las estimaciones realizadas por los diferentes modelos. Se puede observar

que mientras que no se observa una tendencia clara a la infraestimación por parte de MRP, tanto MRQP como MR Negbin II tienden a la sobreestimación de los errores estándar

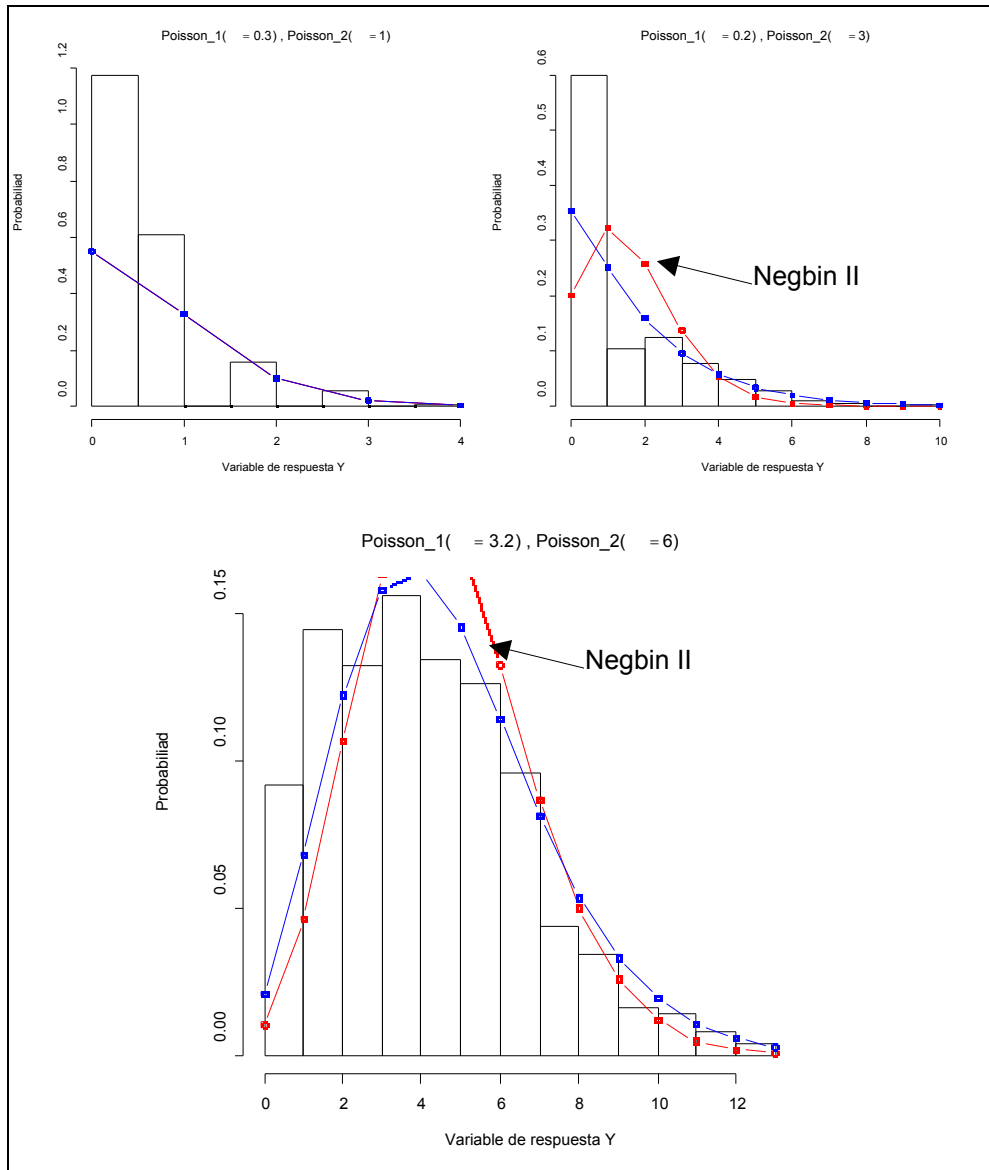


Ilustración 9. Histogramas de sendas muestras de tamaño $n = 500$ generadas como mezcla de dos modelos de Poisson con las distribuciones teóricas Negbin y Poisson solapadas.

Tabla 14. Muestras por mezcla de dos distribuciones de Poisson: potencia estadística de los tests diagnósticos de sobredispersión.

n	λ_1	λ_2	LR			Wald			LMR Negbin I		
			LR	P _{2.5}	P _{97.5}	W	P _{2.5}	P _{97.5}	LMR I	P _{2.5}	P _{97.5}
20	0.2	1	0.043	-0.13	3.64	0.000	0.50	1.20	0.039	-1.13	1.57
20	0.2	3	0.249	-0.04	9.11	0.000	0.64	1.61	0.064	0.35	2.39
20	3.2	6	0.052	-0.11	4.12	0.000	0.50	1.34	0.100	-1.01	1.59
50	0.2	1	0.074	-0.23	4.75	0.000	0.73	1.52	0.053	-0.69	1.71
50	0.2	3	0.517	0.79	15.22	0.000	0.99	2.28	0.247	1.12	2.82
50	3.2	6	0.094	-0.18	5.38	0.000	0.72	1.69	0.053	-0.46	1.89
100	0.2	1	0.110	-0.31	6.09	0.000	0.98	1.80	0.043	-0.33	1.91
100	0.2	3	0.775	3.05	22.12	0.070	1.51	2.92	0.624	1.75	3.45
100	3.2	6	0.147	-0.19	6.83	0.001	0.96	2.02	0.059	-0.13	2.17
500	0.2	1	0.302	-0.22	9.79	0.087	2.10	2.91	0.157	0.49	2.62
500	0.2	3	1.000	26.86	66.94	0.999	4.04	5.55	1.000	4.33	6.06
500	3.2	6	0.493	0.82	12.61	0.072	2.03	3.03	0.307	0.90	3.14
n	λ_1	λ_2	LMR Negbin II			D			χ^2		
			LMR II	P _{2.5}	P _{97.5}	D	P _{2.5}	P _{97.5}	χ^2	P _{2.5}	P _{97.5}
20	0.2	1	0.082	-1.19	1.82	0.024	14.84	28.60	0.066	14.27	34.36
20	0.2	3	0.066	-0.00	2.39	0.314	20.77	41.55	0.312	19.52	45.86
20	3.2	6	0.100	-0.97	1.65	0.148	16.30	37.60	0.104	15.47	35.63
50	0.2	1	0.068	-0.73	1.88	0.020	43.18	64.37	0.096	43.36	75.49
50	0.2	3	0.124	0.65	2.64	0.567	59.48	92.46	0.539	57.76	101.61
50	3.2	6	0.054	-0.48	1.90	0.229	48.54	81.05	0.153	45.87	77.35
100	0.2	1	0.051	-0.39	2.02	0.019	91.72	120.82	0.123	93.41	139.07
100	0.2	3	0.304	1.16	3.09	0.804	125.22	171.25	0.765	123.24	182.33
100	3.2	6	0.062	-0.12	2.18	0.329	103.79	149.26	0.194	98.14	142.04
500	0.2	1	0.124	0.36	2.60	0.021	483.79	548.33	0.249	503.50	602.10
500	0.2	3	0.992	3.13	5.00	1.000	666.37	765.89	1.000	666.73	797.86
500	3.2	6	0.290	0.89	3.09	0.823	559.65	656.83	0.535	530.01	623.25

Tabla 15. Muestras por mezcla de dos distribuciones de Poisson:
estimaciones del parámetro de dispersión.

n	λ_1	λ_2	Negbin II			Quasi-Poisson		
			α	$P_{2.5}$	$P_{97.5}$	ϕ	$P_{2.5}$	$P_{97.5}$
20	0.2	1	15.542	3.471	43.027	1.036	0.788	1.907
20	0.2	3	14.358	2.487	57.768	1.428	1.081	2.541
20	3.2	6	121.824	21.373	386.298	1.126	0.858	1.979
50	0.2	1	10.886	3.501	28.370	1.055	0.895	1.570
50	0.2	3	7.640	2.712	32.735	1.414	1.196	2.116
50	3.2	6	81.695	21.646	233.860	1.119	0.955	1.611
100	0.2	1	9.242	3.942	22.134	1.055	0.944	1.419
100	0.2	3	5.587	2.900	20.351	1.403	1.251	1.860
100	3.2	6	65.989	24.497	178.949	1.114	1.001	1.449
500	0.2	1	7.780	5.287	14.239	1.052	1.000	1.209
500	0.2	3	3.975	3.256	6.475	1.401	1.332	1.602
500	3.2	6	48.447	30.758	104.318	1.112	1.064	1.251

Tabla 16. Muestras por mezcla de dos distribuciones de Poisson: estimaciones de los coeficientes del modelo de regresión para recuentos.

n	λ_1	λ_2	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0.2	1	-0.824	-1.021	-0.207	-0.811	-1.006	-0.2010	-0.811	-1.006	-0.201
20	0.2	3	0.031	-0.113	0.434	0.080	-0.060	0.4737	0.080	-0.060	0.473
20	3.2	6	1.478	1.406	1.678	1.478	1.407	1.6783	1.478	1.407	1.678
50	0.2	1	-0.739	-0.875	-0.347	-0.733	-0.867	-0.3433	-0.733	-0.867	-0.343
50	0.2	3	0.063	-0.028	0.326	0.108	0.016	0.3694	0.108	0.016	0.369
50	3.2	6	1.486	1.441	1.620	1.486	1.441	1.6209	1.486	1.441	1.620
100	0.2	1	-0.719	-0.819	-0.436	-0.714	-0.814	-0.4334	-0.714	-0.814	-0.433
100	0.2	3	0.068	0.002	0.256	0.112	0.048	0.2948	0.112	0.048	0.294
100	3.2	6	1.488	1.455	1.579	1.488	1.455	1.5792	1.488	1.455	1.579
500	0.2	1	-0.705	-0.749	-0.576	-0.702	-0.745	-0.5713	-0.702	-0.745	-0.571
500	0.2	3	0.075	0.045	0.158	0.119	0.090	0.2019	0.119	0.090	0.201
500	3.2	6	1.490	1.475	1.531	1.490	1.475	1.5312	1.490	1.475	1.531
n	λ_1	λ_2	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0.2	1	0.379	0.249	0.800	0.362	0.240	0.754	0.362	0.240	0.754
20	0.2	3	0.542	0.464	0.772	0.483	0.418	0.684	0.483	0.418	0.684
20	3.2	6	0.147	0.108	0.263	0.146	0.108	0.259	0.146	0.108	0.259
50	0.2	1	0.367	0.291	0.598	0.357	0.286	0.578	0.357	0.286	0.578
50	0.2	3	0.548	0.499	0.701	0.492	0.450	0.617	0.492	0.450	0.617
50	3.2	6	0.152	0.126	0.226	0.151	0.125	0.225	0.151	0.125	0.225
100	0.2	1	0.366	0.312	0.524	0.358	0.307	0.510	0.358	0.307	0.510
100	0.2	3	0.551	0.516	0.657	0.495	0.465	0.585	0.495	0.465	0.585
100	3.2	6	0.153	0.135	0.206	0.153	0.135	0.205	0.153	0.135	0.205
500	0.2	1	0.364	0.340	0.433	0.358	0.335	0.424	0.358	0.335	0.424
500	0.2	3	0.554	0.538	0.602	0.497	0.484	0.537	0.497	0.484	0.537
500	3.2	6	0.154	0.145	0.179	0.153	0.145	0.178	0.153	0.145	0.178

Tabla 17. Muestras por mezcla de dos distribuciones de Poisson: estimaciones de los errores estándar de los coeficientes del modelo de regresión para recuentos.

n	λ_1	λ_2	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0.2	1	0.405	0.32	0.71	0.379	0.30	0.64	0.375	0.30	0.61
20	0.2	3	0.287	0.25	0.39	0.248	0.22	0.32	0.289	0.25	0.40
20	3.2	6	0.116	0.10	0.14	0.108	0.10	0.12	0.114	0.10	0.15
50	0.2	1	0.235	0.21	0.31	0.223	0.20	0.30	0.227	0.20	0.31
50	0.2	3	0.178	0.16	0.21	0.153	0.14	0.18	0.180	0.16	0.22
50	3.2	6	0.072	0.06	0.08	0.068	0.06	0.07	0.072	0.06	0.08
100	0.2	1	0.162	0.15	0.19	0.155	0.14	0.19	0.158	0.14	0.19
100	0.2	3	0.125	0.11	0.14	0.108	0.10	0.12	0.127	0.12	0.14
100	3.2	6	0.051	0.04	0.05	0.048	0.04	0.05	0.050	0.04	0.05
500	0.2	1	0.071	0.06	0.07	0.068	0.06	0.07	0.070	0.06	0.07
500	0.2	3	0.055	0.05	0.05	0.048	0.04	0.05	0.056	0.05	0.06
500	3.2	6	0.022	0.02	0.02	0.021	0.02	0.02	0.022	0.02	0.02
n	λ_1	λ_2	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0.2	1	0.207	0.17	0.34	0.191	0.16	0.30	0.189	0.16	0.30
20	0.2	3	0.147	0.12	0.20	0.121	0.11	0.15	0.142	0.12	0.19
20	3.2	6	0.063	0.05	0.08	0.058	0.05	0.06	0.061	0.05	0.08
50	0.2	1	0.125	0.11	0.16	0.117	0.10	0.15	0.120	0.10	0.15
50	0.2	3	0.094	0.08	0.11	0.077	0.07	0.09	0.091	0.08	0.11
50	3.2	6	0.040	0.03	0.04	0.038	0.03	0.04	0.040	0.03	0.04
100	0.2	1	0.088	0.08	0.10	0.082	0.07	0.09	0.084	0.07	0.10
100	0.2	3	0.067	0.06	0.07	0.055	0.05	0.06	0.065	0.06	0.07
100	3.2	6	0.028	0.02	0.03	0.027	0.02	0.02	0.028	0.02	0.03
500	0.2	1	0.039	0.03	0.04	0.037	0.03	0.03	0.037	0.03	0.04
500	0.2	3	0.030	0.02	0.03	0.024	0.02	0.02	0.029	0.02	0.03
500	3.2	6	0.012	0.01	0.01	0.012	0.01	0.01	0.012	0.01	0.01

Tabla 18. Muestras por mezcla de dos distribuciones de Poisson: estimaciones Monte Carlo de los errores estándar de los coeficientes de regresión.

n	λ_1	λ_2	b_0			b_1		
			MR Negbin II	MRP	MRQP	MR Negbin II	MRP	MRQP
20	0.2	1	0.439	0.427	0.427	0.200	0.188	0.188
20	0.2	3	0.230	0.226	0.226	0.113	0.095	0.095
20	3.2	6	0.107	0.107	0.107	0.057	0.057	0.057
50	0.2	1	0.229	0.228	0.228	0.112	0.108	0.108
50	0.2	3	0.143	0.141	0.141	0.072	0.062	0.062
50	3.2	6	0.068	0.068	0.068	0.038	0.037	0.037
100	0.2	1	0.154	0.153	0.153	0.078	0.075	0.075
100	0.2	3	0.103	0.100	0.100	0.052	0.044	0.044
100	3.2	6	0.048	0.048	0.048	0.026	0.026	0.026
500	0.2	1	0.067	0.067	0.067	0.034	0.033	0.033
500	0.2	3	0.044	0.043	0.043	0.023	0.019	0.019
500	3.2	6	0.021	0.021	0.021	0.012	0.012	0.012

5.2.3 Heterogeneidad observada simulada mediante un modelo ZIP (exceso de ceros)

Exceso de ceros (modelo ZIP –heterogeneidad observada–) mediante la mezcla de dos mecanismos generadores: un modelo logit para predecir si $Y = 0$ o bien, si $Y \sim \text{Poisson}(\lambda)$, y un modelo de Poisson para el segundo caso. Por tanto, en este caso el mecanismo generador maneja dos grupos de parámetros (Long, 1997): β_0 y β_1 para la parte *logit*, y γ_0 y γ_1 para la parte Poisson. Concretamente se generarán 5,000 muestras con los siguientes patrones muestrales:

β_0	β_1	γ_0	γ_1
0	1	0	-0.5
0	1	0	0.5
0	1	0	1.25

para cada uno de los tamaños muestrales utilizados:

$$n = 500, n = 100, n = 50 \text{ y } n = 20$$

1. *Potencia estadística de las pruebas diagnósticas de sobredispersión:* Se observa que, para todas las pruebas, y al igual que en los experimentos anteriores, la potencia mejora a medida que aumenta la sobredispersión y el tamaño muestral (ver Tabla 19).
2. *Estimaciones de los parámetros de dispersión:* Con tamaños muestrales $n = 20$ y $n = 50$ las estimaciones de los parámetros de dispersión son menores a las esperadas, tanto de Negbin II como mediante quasi-Poisson (ver Tabla 20). Para los tamaños muestrales $n = 100$ y $n = 500$ tanto como se aproximan bien al valor esperado.
3. *Estabilidad de las estimaciones de β_0 y β_1 a lo largo de los diferentes modelos de regresión ajustados:* la Tabla 21 muestra que los valores de β_0 y β_1 obtenidos mediante MRP, MRQP y MR Negbin II son muy parecidos. Al igual que ocurría en el primero de este grupo de estudios, la convergencia mutua de los valores de los parámetros se hace más evidente a medida que la muestra aumenta, hasta llegar a $n = 500$ en que dichos valores son casi idénticos.

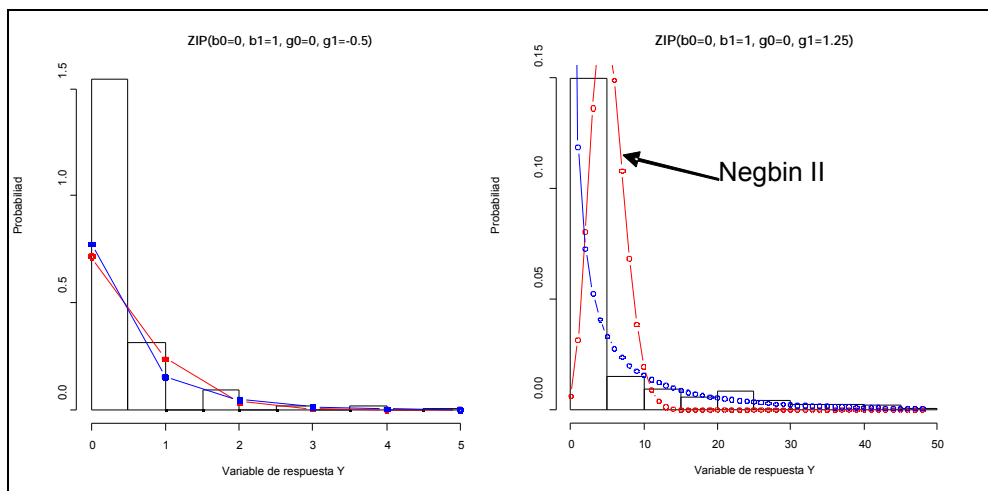


Ilustración 10. Histogramas de sendas muestras de tamaño $n = 500$ generadas a partir de dos modelos ZIP con las distribuciones teóricas Negbin y Poisson solapadas.

4. *Estimaciones de los errores estándar de los coeficientes de regresión:* En la Tabla 22 se presentan las estimaciones de los errores estándar de los coeficientes de regresión; en la Tabla 23 se presentan las estimaciones Monte Carlo de los verdaderos errores estándar que sirven de referencia para valorar las estimaciones realizadas por los diferentes modelos. Se puede observar que tanto MRP como MRQP infraestiman los errores estándar, aunque esta diferencia entre errores estándar estimados y verdaderos se diluye a medida

que el tamaño de la muestra aumenta. Por otro lado, las estimaciones con MR Negbin II funcionan correctamente para muestras moderadamente grandes ($n = 100$ y $n = 500$).

Tabla 19. Muestras con exceso de ceros:
potencia estadística de los tests diagnósticos de sobredispersión.

n	γ_1	γ_2	LR			Wald			LMR Negbin I		
			LR	P _{2.5}	P _{97.5}	W	P _{2.5}	P _{97.5}	LMR I	P _{2.5}	P _{97.5}
20	0	-0.5	0.360	0.004	13.108	0.000	0.477	1.379	0.048	0.264	2.369
20	0	0.5	0.118	-0.174	6.389	0.000	0.617	1.494	0.046	-0.302	2.113
20	0	1.25	0.325	-0.454	47.586	0.008	0.946	2.181	0.058	0.271	2.244
50	0	-0.5	0.740	2.561	26.468	0.000	1.069	2.038	0.175	1.148	2.730
50	0	0.5	0.261	-0.132	10.036	0.007	0.940	2.143	0.108	0.643	2.454
50	0	1.25	0.584	-0.057	87.510	0.055	1.380	3.442	0.277	1.227	2.802
100	0	-0.5	0.943	8.987	43.773	0.035	1.789	2.744	0.463	1.612	3.072
100	0	0.5	0.446	0.389	14.245	0.026	1.356	2.720	0.336	1.250	2.991
100	0	1.25	0.798	4.742	126.777	0.401	2.051	4.983	0.692	1.876	3.369
500	0	-0.5	1.000	71.003	147.541	0.999	4.583	5.577	0.998	3.351	5.086
500	0	0.5	0.961	9.203	36.800	0.986	3.152	4.643	0.994	3.294	5.020
500	0	1.25	0.996	102.555	347.388	1.000	5.064	9.901	0.999	4.509	6.018
n	γ_1	γ_2	LMR Negbin II			D			χ^2		
			LMR II	P _{2.5}	P _{97.5}	D	P _{2.5}	P _{97.5}	χ^2	P _{2.5}	P _{97.5}
20	0	-0.5	0.137	0.139	3.349	0.176	14.72	38.90	0.444	19.46	55.86
20	0	0.5	0.131	-0.796	2.891	0.064	13.85	31.81	0.132	13.92	39.50
20	0	1.25	0.188	-0.751	7.565	0.358	15.66	86.72	0.301	13.45	72.51
50	0	-0.5	0.264	1.053	3.317	0.291	47.61	87.26	0.811	69.03	137.62
50	0	0.5	0.116	0.046	2.891	0.085	43.26	71.84	0.240	45.24	93.05
50	0	1.25	0.290	0.208	6.224	0.588	53.63	166.09	0.459	47.61	135.74
100	0	-0.5	0.472	1.500	3.565	0.436	105.44	161.35	0.967	158.08	262.42
100	0	0.5	0.181	0.454	3.214	0.103	93.21	133.64	0.348	100.50	171.71
100	0	1.25	0.500	0.985	6.370	0.791	127.73	280.57	0.620	111.00	226.93
500	0	-0.5	0.992	3.255	5.457	0.899	579.06	705.58	1.000	903.15	1138.63
500	0	0.5	0.681	1.790	4.412	0.243	506.42	593.99	0.837	565.52	720.06
500	0	1.25	0.985	4.362	9.145	0.999	763.32	1087.23	0.972	642.83	880.42

Tabla 20. Muestras con exceso de ceros:
estimaciones del parámetro de dispersión.

n	γ_0	γ_1	Negbin II			Quasi-Poisson		
			α	$P_{2.5}$	$P_{97.5}$	ϕ	$P_{2.5}$	$P_{97.5}$
20	0	-0.5	4.377	0.319	23.974	1.585	1.079	3.083
20	0	0.5	17.986	2.821	61.280	1.103	0.769	2.188
20	0	1.25	89.874	1.946	161.119	1.433	0.746	4.012
50	0	-0.5	1.039	0.299	7.919	1.774	1.436	2.838
50	0	0.5	9.101	2.738	32.729	1.168	0.936	1.935
50	0	1.25	19.509	1.907	95.209	1.412	0.989	2.816
100	0	-0.5	0.537	0.309	1.631	1.859	1.609	2.652
100	0	0.5	6.614	2.931	20.375	1.183	1.016	1.750
100	0	1.25	9.760	2.078	57.535	1.403	1.129	2.299
500	0	-0.5	0.394	0.337	0.585	1.921	1.806	2.263
500	0	0.5	4.670	3.474	8.706	1.201	1.123	1.445
500	0	1.25	3.797	2.490	9.603	1.406	1.287	1.760

Tabla 21. Muestras con exceso de ceros:
estimaciones de los coeficientes
del modelo de regresión para recuentos.

n	γ_1	γ_2	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0	-0.5	-1.625	-1.668	-0.320	-1.455	-1.652	-0.301	-1.455	-1.652	-0.301
20	0	0.5	-0.963	-1.228	-0.132	-0.921	-1.184	-0.113	-0.921	-1.184	-0.113
20	0	1.25	-0.665	-0.924	0.126	-0.598	-0.868	0.229	-0.598	-0.868	0.229
50	0	-0.5	-1.106	-1.312	-0.512	-1.100	-1.304	-0.512	-1.100	-1.304	-0.512
50	0	0.5	-0.844	-1.013	-0.323	-0.813	-0.983	-0.305	-0.813	-0.983	-0.305
50	0	1.25	-0.652	-0.841	-0.139	-0.558	-0.739	-0.038	-0.558	-0.739	-0.038
100	0	-0.5	-1.032	-1.175	-0.621	-1.029	-1.173	-0.619	-1.029	-1.173	-0.619
100	0	0.5	-0.818	-0.949	-0.452	-0.789	-0.916	-0.428	-0.789	-0.916	-0.428
100	0	1.25	-0.654	-0.791	-0.262	-0.551	-0.686	-0.167	-0.551	-0.686	-0.167
500	0	-0.5	-0.989	-1.058	-0.800	-0.988	-1.058	-0.799	-0.988	-1.058	-0.799
500	0	0.5	-0.797	-0.858	-0.626	-0.770	-0.831	-0.603	-0.770	-0.831	-0.603
500	0	1.25	-0.650	-0.713	-0.468	-0.543	-0.604	-0.365	-0.543	-0.604	-0.365
n	γ_1	γ_2	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0	-0.5	0.134	-0.147	0.893	0.093	-0.120	0.788	0.093	-0.120	0.788
20	0	0.5	0.858	0.700	1.415	0.826	0.673	1.348	0.826	0.673	1.348
20	0	1.25	1.479	1.340	1.958	1.439	1.322	1.881	1.439	1.322	1.881
50	0	-0.5	0.037	-0.107	0.456	0.029	-0.092	0.414	0.029	-0.092	0.414
50	0	0.5	0.826	0.736	1.126	0.799	0.709	1.093	0.799	0.709	1.093
50	0	1.25	1.487	1.399	1.769	1.433	1.356	1.681	1.433	1.356	1.681
100	0	-0.5	0.019	-0.081	0.311	0.015	-0.072	0.277	0.015	-0.072	0.277
100	0	0.5	0.819	0.753	1.019	0.795	0.729	0.986	0.795	0.729	0.986
100	0	1.25	1.493	1.429	1.690	1.434	1.377	1.612	1.434	1.377	1.612
500	0	-0.5	0.003	-0.042	0.132	0.002	-0.038	0.116	0.002	-0.038	0.116
500	0	0.5	0.815	0.787	0.900	0.792	0.764	0.874	0.792	0.764	0.874
500	0	1.25	1.493	1.465	1.578	1.432	1.405	1.509	1.432	1.405	1.509

Tabla 22. Muestras con exceso de ceros:
estimaciones de los errores estándar de los coeficientes
del modelo de regresión para recuentos.

n	γ_1	γ_2	MR Negbin II			MRP			MRQP		
			b_0	$P_{2.5}$	$P_{97.5}$	B_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0	-0.5	20.742	0.454	1.555	0.982	0.348	1.414	0.575	0.443	1.128
20	0	0.5	0.499	0.388	0.941	0.457	0.358	0.835	0.457	0.362	0.794
20	0	1.25	0.447	0.340	0.829	0.348	0.305	0.498	0.399	0.282	0.841
50	0	-0.5	0.342	0.295	0.504	0.256	0.218	0.402	0.333	0.290	0.480
50	0	0.5	0.290	0.254	0.420	0.267	0.234	0.383	0.284	0.245	0.423
50	0	1.25	0.291	0.239	0.444	0.219	0.202	0.270	0.255	0.212	0.395
100	0	-0.5	0.233	0.213	0.301	0.170	0.154	0.223	0.230	0.210	0.295
100	0	0.5	0.201	0.183	0.259	0.185	0.169	0.235	0.200	0.181	0.261
100	0	1.25	0.208	0.183	0.284	0.155	0.146	0.179	0.182	0.161	0.250
500	0	-0.5	0.103	0.099	0.115	0.073	0.070	0.081	0.101	0.097	0.113
500	0	0.5	0.088	0.085	0.099	0.081	0.078	0.091	0.089	0.085	0.101
500	0	1.25	0.094	0.089	0.108	0.069	0.067	0.074	0.082	0.078	0.094
n	γ_1	γ_2	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0	-0.5	7.602	0.245	0.776	0.493	0.185	0.669	0.297	0.237	0.565
20	0	0.5	0.230	0.183	0.404	0.200	0.164	0.334	0.201	0.163	0.333
20	0	1.25	0.204	0.139	0.415	0.136	0.122	0.188	0.156	0.111	0.320
50	0	-0.5	0.192	0.166	0.280	0.141	0.122	0.214	0.184	0.162	0.262
50	0	0.5	0.141	0.125	0.199	0.122	0.110	0.167	0.130	0.114	0.187
50	0	1.25	0.140	0.103	0.233	0.088	0.083	0.106	0.103	0.086	0.159
100	0	-0.5	0.133	0.121	0.172	0.096	0.088	0.125	0.130	0.119	0.165
100	0	0.5	0.099	0.091	0.127	0.086	0.080	0.105	0.093	0.085	0.118
100	0	1.25	0.102	0.085	0.148	0.063	0.060	0.072	0.074	0.066	0.102
500	0	-0.5	0.059	0.057	0.066	0.042	0.040	0.046	0.058	0.056	0.065
500	0	0.5	0.044	0.042	0.049	0.038	0.037	0.042	0.042	0.040	0.047
500	0	1.25	0.047	0.044	0.056	0.028	0.028	0.030	0.034	0.032	0.039

Tabla 23. Muestras con exceso de ceros:
estimaciones Monte Carlo de los errores estándar
de los coeficientes del modelo de regresión para recuentos.

n	γ_1	γ_2	b_0			b_1		
			MR Negbin II	MRP	MRQP	MR Negbin II	MRP	MRQP
20	0	-0.5	5.642	2.236	2.236	1.882	0.749	0.749
20	0	0.5	0.592	0.573	0.573	0.237	0.229	0.229
20	0	1.25	0.527	0.513	0.513	0.213	0.209	0.209
50	0	-0.5	0.371	0.368	0.368	0.212	0.192	0.192
50	0	0.5	0.313	0.309	0.309	0.135	0.133	0.133
50	0	1.25	0.311	0.295	0.295	0.128	0.124	0.124
100	0	-0.5	0.238	0.237	0.237	0.145	0.131	0.131
100	0	0.5	0.213	0.209	0.209	0.094	0.092	0.092
100	0	1.25	0.222	0.211	0.211	0.091	0.088	0.088
500	0	-0.5	0.103	0.103	0.103	0.064	0.058	0.058
500	0	0.5	0.093	0.092	0.092	0.041	0.041	0.041
500	0	1.25	0.098	0.093	0.093	0.041	0.039	0.039

5.3 Comparación de procedimientos para la corrección del error estándar de las estimaciones de los coeficientes del MRP en presencia de sobredispersión

Para cada muestra generada en el experimento 2 con el generador Negbin II, se comparan los siguientes procedimientos para la corrección del EE de los coeficientes de regresión MRP:

- EE estimado por MRP y multiplicado por $\sqrt{\sigma^2 / gl}$
- EE estimado por MRP y multiplicado por $\sqrt{D / gl}$
- EE estimado por MRP y multiplicado por parámetro de dispersión estimado, por ejemplo, por quasi-verosimilitud (quasi-Poisson)
- Estimación bootstrap no paramétrica del EE
- Estimación jackknife del EE

Para este estudio, se extraen 3,000 muestras cuyos tamaños son únicamente de $n = 20$ y $n = 100$, puesto que el objetivo es comparar muestras de tamaño medio con muestras pequeñas.

Los resultados de las estimaciones mostrados en la Tabla 24 y la Tabla 25 muestran en general que los valores de las estimaciones del error estándar dependen no tan sólo del tamaño muestral sino también del parámetro de dispersión (σ).

Cabe destacar que las correcciones del error estándar que mejor se ajustan al valor verdadero del error estándar son las basadas en Jackknife y Bootstrap. Lo especialmente destacable de ambos procedimientos es que tal efectividad en las correcciones se da, tal como indican Cameron y Trivedi (Cameron y Trivedi, 1998), incluso en muestras pequeñas y parámetros de dispersión pequeños. En este sentido, las estimaciones basadas en los dos métodos anteriores son mejores que las basadas en discrepancia, σ^2 y D , los cuales se caracterizan por una infraestimación del error estándar, sobretodo en muestras pequeñas. Tal infraestimación es más pronunciada en la corrección basada en la discrepancia, que en las basadas en σ^2 y D , siendo estas dos últimas casi idénticas en los valores de corrección que proporcionan.

Tabla 24. Muestras Negbin II:
comparación de diferentes correcciones
del error estándar del coeficiente b_0 .

n	μ	τ	Monte Carlo	EE(b_0) Jackknife			EE(b_0) Bootstrap					
			b_0	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$			
20	0.3	0.15	6.467	5.330	0.723	47.878	6.001	2.200	20.839			
20	0.3	0.3	4.480	4.006	0.597	47.532	5.402	1.422	20.040			
20	1	0.5	0.450	0.487	0.361	0.978	0.967	0.369	6.051			
20	1	1	0.359	0.378	0.297	0.685	0.551	0.306	3.616			
20	5	2.5	0.179	0.181	0.156	0.262	0.173	0.150	0.248			
20	5	5	0.143	0.146	0.128	0.205	0.140	0.122	0.197			
100	0.3	0.15	0.450	0.363	0.304	0.587	0.377	0.305	0.606			
100	0.3	0.3	0.279	0.286	0.246	0.422	0.288	0.245	0.427			
100	1	0.5	0.229	0.180	0.163	0.237	0.175	0.160	0.226			
100	1	1	0.147	0.146	0.135	0.183	0.144	0.133	0.180			
100	5	2.5	0.100	0.078	0.074	0.092	0.077	0.072	0.093			
100	5	5	0.064	0.064	0.061	0.074	0.063	0.059	0.075			
n	μ	τ	Monte Carlo	Discrepancia $\sqrt{(D/gI)} \times EE_{MRP}(b_0)$			χ^2 $\sqrt{(\chi^2/gI)} \times EE_{MRP}(b_0)$			ϕ Quasi-Poisson $\sqrt{\phi} \times EE_{MRP}(b_0)$		
			b_0	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$	b_0	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	6.467	0.614	0.454	1.911	0.782	0.556	1.818	0.758	0.544	1.816
20	0.3	0.3	4.480	0.552	0.425	1.482	0.676	0.476	1.618	0.667	0.474	1.601
20	1	0.5	0.450	0.362	0.322	0.513	0.390	0.324	0.621	0.390	0.324	0.620
20	1	1	0.359	0.319	0.285	0.442	0.324	0.273	0.509	0.323	0.272	0.508
20	5	2.5	0.179	0.172	0.154	0.226	0.170	0.149	0.230	0.170	0.149	0.230
20	5	5	0.143	0.142	0.126	0.192	0.139	0.123	0.191	0.139	0.123	0.191
100	0.3	0.15	0.450	0.217	0.204	0.270	0.318	0.279	0.456	0.317	0.279	0.454
100	0.3	0.3	0.279	0.199	0.188	0.237	0.262	0.232	0.362	0.261	0.232	0.360
100	1	0.5	0.229	0.152	0.146	0.172	0.172	0.158	0.219	0.172	0.158	0.218
100	1	1	0.147	0.135	0.130	0.152	0.142	0.131	0.173	0.142	0.131	0.172
100	5	2.5	0.100	0.077	0.074	0.087	0.077	0.073	0.090	0.077	0.073	0.090
100	5	5	0.064	0.064	0.061	0.072	0.063	0.060	0.073	0.063	0.060	0.073

Tabla 25. Muestras Negbin II:
comparación de diferentes correcciones
del error estándar del coeficiente b_1 .

n	μ	τ	Monte Carlo	EE(b_1) Jackknife			EE(b_1) Bootstrap					
			b_1	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$			
20	0.3	0.15	2.318	1.700	0.257	17.481	2.079	0.286	7.782			
20	0.3	0.3	1.281	1.374	0.248	16.389	1.930	0.298	7.285			
20	1	0.5	0.229	0.246	0.171	0.554	0.402	0.180	2.231			
20	1	1	0.193	0.197	0.145	0.410	0.253	0.151	1.334			
20	5	2.5	0.096	0.099	0.081	0.166	0.094	0.080	0.144			
20	5	5	0.081	0.080	0.066	0.126	0.076	0.064	0.116			
100	0.3	0.15	0.195	0.195	0.149	0.354	0.192	0.147	0.342			
100	0.3	0.3	0.155	0.155	0.125	0.269	0.152	0.124	0.255			
100	1	0.5	0.100	0.100	0.085	0.150	0.096	0.083	0.138			
100	1	1	0.081	0.083	0.074	0.117	0.081	0.072	0.112			
100	5	2.5	0.045	0.045	0.041	0.057	0.044	0.040	0.058			
100	5	5	0.037	0.036	0.034	0.045	0.036	0.033	0.045			
n	μ	τ	Monte Carlo	Discrepancia $\sqrt{(D/gI) \times EE_{MRP}(b_1)}$			χ^2 $\sqrt{(\chi^2/gI) \times EE_{MRP}(b_1)}$			ϕ Quasi-Poisson $\sqrt{\phi \times EE_{MRP}(b_1)}$		
			b_1	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$	b_1	$P_{2.5}$	$P_{97.5}$
20	0.3	0.15	2.318	0.297	0.243	0.638	0.382	0.292	0.731	0.373	0.285	0.730
20	0.3	0.3	1.281	0.274	0.227	0.577	0.337	0.256	0.648	0.333	0.254	0.647
20	1	0.5	0.229	0.193	0.175	0.255	0.208	0.174	0.320	0.207	0.174	0.319
20	1	1	0.193	0.171	0.155	0.225	0.174	0.148	0.261	0.173	0.147	0.261
20	5	2.5	0.096	0.094	0.084	0.123	0.093	0.082	0.126	0.093	0.082	0.126
20	5	5	0.081	0.078	0.069	0.105	0.076	0.067	0.104	0.076	0.067	0.104
100	0.3	0.15	0.195	0.121	0.116	0.142	0.178	0.158	0.249	0.177	0.157	0.248
100	0.3	0.3	0.155	0.112	0.107	0.129	0.147	0.131	0.203	0.147	0.131	0.203
100	1	0.5	0.100	0.086	0.083	0.097	0.098	0.090	0.123	0.098	0.090	0.123
100	1	1	0.081	0.077	0.074	0.086	0.081	0.075	0.098	0.081	0.075	0.098
100	5	2.5	0.045	0.044	0.042	0.050	0.044	0.042	0.051	0.044	0.042	0.051
100	5	5	0.037	0.037	0.035	0.041	0.036	0.034	0.042	0.036	0.034	0.042

6

Conclusiones

Este trabajo se enmarca en una incipiente línea de trabajo con una clara vocación aplicada: investigar en las condiciones de uso de las técnicas de modelado estadístico específicas y no específicas para datos de recuento, así como ampliar y divulgar las herramientas informáticas que requiere un investigador para el análisis de datos de recuento. De acuerdo con este punto de vista eminentemente práctico, este trabajo no empezó a tomar forma hasta que se evaluó la necesidad de tal entorno de análisis. Esto implicó el estudio de la frecuencia de uso de variables de recuento y del tipo de análisis que se aplica a las mismas.

Los recuentos son un tipo de variables que frecuentemente son objeto de estudio en ámbitos como las Ciencias Sociales o las Ciencias de la Salud. Tal como se ha podido comprobar a través del primer objetivo planteado en el estudio bibliométrico, la Psicología no es una excepción: las variables de recuento tienen una presencia importante en nuestra disciplina, especialmente en ciertos ámbitos aplicados de la misma.

Tal como se ha indicado en repetidas ocasiones en este y otros trabajos, el conocimiento acerca de la naturaleza de una variable así como de sus características distribucionales constituyen la base a partir de la cual se justifica la aplicación de un modelo estadístico determinado. En este sentido, la naturaleza entera y no negativa de una variable de recuento, hace que requiera modelos estadísticos específicos para su análisis. En esta misma línea, el segundo objetivo planteado en el marco del estudio bibliométrico, hacía referencia al estudio de la frecuencia de uso de modelos estadísticos adecuados para el análisis de variables de recuento. Los resultados no dejan lugar a duda: no se ha podido detectar en ninguna investigación la aplicación de un modelo específico para datos de recuento. Siguiendo la estela de muchas otras disciplinas (Ciencias Políticas, Ciencias Económicas, Medicina, etc.), en Psicología el análisis de datos aplicado con mayor frecuencia en estudios con variables de recuento es el análisis de la variancia o la regresión lineal.

Tal como se apuntaba en el estudio bibliométrico, el uso del modelo lineal general como herramienta de análisis multipropósito puede responder a diferentes motivos como, por ejemplo:

- el desconocimiento de la naturaleza idiosincrásica de las variables de recuento,
- el desconocimiento de las implicaciones asociadas al uso de un modelo estadístico inadecuado como el modelo lineal general,

- el desconocimiento de modelos específicos para este tipo de variables,
- la creencia de que los modelos estadísticos específicos para las variables de recuento están todavía en una fase de desarrollo teórico y, por tanto, no son todavía utilizables en la práctica,
- o también, la escasa accesibilidad a programas informáticos para análisis de datos que incorporen dichos modelos.

Si es el desconocimiento el que puede explicar la falta de aplicación de modelos estadísticos adecuados, entonces una tarea importante a realizar es la de aportar la información necesaria para que, al menos, el investigador sea consciente de que está aplicando un modelo de análisis de datos inadecuado y de las consecuencias que ello comporta. En este sentido, el contexto en el cual nos encontramos, el universitario, debe ser sin duda el principal medio de transmisión de esta información; especialmente para las futuras generaciones de investigadores aplicados.

Tal como se ha indicado en el primer capítulo, los cimientos de este trabajo se encuentran en el modelado estadístico como procedimiento de análisis de datos y en el modelo lineal generalizado como marco teórico estadístico. El objetivo de la aplicación conjunta de ambos aspectos es poder disponer de las herramientas de análisis adecuadas para cada tipo de datos y que tal análisis se lleve a cabo siguiendo un procedimiento según las características enunciadas en el primer capítulo. Se trata, en definitiva, de ajustar el modelo a los datos.

De acuerdo con esta vocación de ajuste del modelo a los datos, se ha presentado el que es considerado el modelo de referencia en el análisis de datos de recuento: el modelo de regresión de Poisson. El hecho de que el modelo de regresión de Poisson reconozca la naturaleza de las variables de recuento hace de él un candidato idóneo para el análisis de este tipo de variables, al menos *a priori*. Sin embargo, tal como se ha expuesto, la propia restrictividad del modelo como consecuencia de las asunciones de las que parte, hacen que su aplicabilidad se vea seriamente restringida a, básicamente, una situación de equidispersión. Situación esta que, tal como señalan diversos autores (Long, 1997; Winkelmann, 2000), es considerada más la excepción que la norma. Asimismo, en ausencia de equidispersión, la situación resultante más habitual es la sobredispersión.

La evaluación de la presencia de sobredispersión requiere, habitualmente, de la aplicación de pruebas para su diagnóstico. De las pruebas disponibles para el diagnóstico de la sobredispersión, se han seleccionado aquellas que se aplican con mayor frecuencia (razón de verosimilitud, prueba de Wald, prueba multiplicador de Lagrange (basada en Negbin I y Negbin II), χ^2/gl y D/gl), y se han llevado a cabo diferentes experimentos de simulación con el objetivo de estudiar su tasa nominal de error y su potencia bajo diferentes tamaños muestrales y diferentes tipos y grados de sobredispersión. Nuestros resultados muestran que el único factor que parece incidir claramente en la tasa nominal de error es el tamaño de la muestra. En cuanto a la potencia, ésta depende tanto del tamaño de la muestra como del tamaño del efecto, en este caso la

sobredispersión, de forma que la proporción de decisiones de significación estadística aumenta a medida que lo hace el tamaño muestral y la magnitud de sobredispersión. Por tanto, estos resultados muestran que los tests analizados se comportan coherentemente con la teoría estadística asintótica y paramétrica que los sustenta.

En cuanto al segundo objetivo, los datos van en la misma dirección de lo ya observado por otros autores (Gourieroux et al., 1984a; Long, 1997): en presencia de sobredispersión las estimaciones de los coeficientes en el MRP son insesgadas, aunque las estimaciones de los errores estándar sí presentan un sesgo hacia la infravaloración, bajo cualquier mecanismo generador de sobredispersión, aunque de nuevo, el tamaño muestral y el grado de sobredispersión inciden directamente sobre el nivel de infraestimación de la dispersión de los parámetros del modelo de regresión.

Por último, en cuanto a los procedimientos de corrección y estimación del error estándar de los coeficientes de regresión por Poisson, claramente los resultados indican la superioridad de las estimaciones no paramétricas *bootstrap* (y también *jackknife*, aunque ligeramente menos eficientes), sobre la corrección directa del error estándar infraestimado mediante su producto por la raíz de alguna forma de estimación del parámetro de dispersión (procedimientos éstos que, como hemos visto, no son válidos para tamaños muestrales $n = 20$, pero sí para $n = 100$).

En general, respecto a los resultados de las simulaciones, las líneas de trabajo más inmediatas son:

- Ampliar las condiciones bajo las que se llevan a cabo las mismas, para poder estudiar el comportamiento de las pruebas de diagnóstico bajo un mayor número de situaciones.
- Ampliar el número de pruebas de detección de sobredispersión, incluyendo las pruebas para modelos no anidados. De todas formas, cabe recordar que, tal como indica Winkelmann (2000), muchas de estas pruebas están aún en fase de desarrollo.
- Profundizar en la aplicación de las estimaciones *bootstrap* del error estándar de los coeficientes del MRP, especialmente sus versiones paramétricas y semi-paramétricas, en la línea apuntada por autores como Cameron y Trivedi (1998).

El objetivo de tales ampliaciones sería ir construyendo el criterio de selección de una u otra prueba diagnóstica, en función de aspectos como el tamaño de la muestra, el valor esperado, o el valor de algún indicador de dispersión.

Tal como se indicaba al principio de este apartado, el objetivo que aparece en el horizonte es el desarrollo de un entorno integrado y coherente de herramientas para el análisis de datos de recuento. En este sentido, cabe recordar que, en general, el investigador aplicado concibe el análisis estadístico como un medio a través del cual poder obtener resultados que permitan derivar conclusiones. De esta forma, resulta imprescindible que los procedimientos de análisis de datos

sean aplicables en un tiempo razonable. Ello implica que el procedimiento debe estar disponible y en un entorno que facilite su implementación o su parametrización. El entorno al que nos referimos es, lógicamente, un entorno de software para el análisis estadístico. Sin embargo, el problema que presenta el software estadístico para datos de recuento quizás sea el hecho de que no exista un paquete que ponga al alcance del investigador la «caja de herramientas» para el modelado de recuentos, sino que, habitualmente estas herramientas se presentan sin caja, esto es, algo dispersas, e incompletas. En este sentido, uno de los objetivos que nos planteamos para el futuro es la implementación de esta «caja de herramientas» en un entorno intuitivo y orientado al modelado, es decir, coherente en la forma en que se indican los modelos y en la información que éstos retornan para su interpretación; sin duda el enfoque orientado a objetos que implementan entornos modernos como R y S-PLUS son el referente a seguir y a potenciar.

Anexos

Listado 1. Programa R para la Simulación 1.

```
# ===== #
# SIMULACIÓN 1: Muestras Poisson #
# #
# Estudio comparativo de la tasa nominal de error de los #
# tests diagnósticos de sobredispersión de uso habitual #
# ===== #

# Carga de funciones para el ajuste de modelos y
# tests de sobredispersión.
source("c:/simula/modelos.r")
source("c:/simula/tests.r")
library( MASS )

# Inicialización de parámetros de la simulación
set.seed(30)
tpo <- Sys.time()
n <- c( 500, 100, 50, 20 )
nN <- length( n )
mu <- c( 0.5, 1, 5, 10 )
nB <- length(mu)
nSim <- 5000
rEqui <- matrix(NA,nN*nB*nSim,14)
rcto <- 0
plot.new()

# Simulación
for (i in 1:nN)
{
  # Generación de la variable independiente X
  x <- seq( from=-3, to=3, length=n[i] )

  for (j in 1:nB)
  {
    for (k in 1:nSim)
    {
      # Contador de muestras generadas
      rcto <- rcto + 1
      title(rcto,col.main="blue")

      # Generación de la variable de respuesta Y bajo
      # distribución de Poisson
      y <- rpois( n[i],lambda=mu[j] )

      # Ajuste de modelos para recuentos (MRP y Negbin II)
      poi <- glm.poisson( y, x )
      nb <- glm.negbin( y, x )
    }
  }
}
```

```

# Tests de sobredispersión
o.lr   <- LR( poi$LL, nb$LL )
o.w    <- W( nb$modelo$theta, nb$modelo$SE.theta )
o.lm   <- LM( y, poi$modelo$fit )
o.lmr  <- LMR( y, poi$modelo$fit )

# Almacena vector de resultados
rEqui[rcto,1] <- n[i]
rEqui[rcto,2] <- mu[j]
rEqui[rcto,3] <- o.lr$val
rEqui[rcto,4] <- o.lr$sig
rEqui[rcto,5] <- o.w$val
rEqui[rcto,6] <- o.w$sig
rEqui[rcto,7] <- o.lmr$val[1]
rEqui[rcto,8] <- o.lmr$sig[1]
rEqui[rcto,9] <- o.lmr$val[2]
rEqui[rcto,10] <- o.lmr$sig[2]
rEqui[rcto,11] <- poi$modelo$dev
rEqui[rcto,12] <- (1 - pchisq( poi$modelo$dev,
                             poi$modelo$df.res )) <= 0.05
rEqui[rcto,13] <- poi$X2
rEqui[rcto,14] <- (1 - pchisq( poi$X2,
                             poi$modelo$df.res )) <= 0.05

# Borra número de iteración
title(rcto,col.main="white")
}
}
}
print( difftime(Sys.time(),tpo) )

# Almacena en disco las matrices de resultados para cada n
write(t(rEqui[1:(nN*nSim*1)],),
      "c:/simula/sim1_n500.res", ncolumns=14)
write(t(rEqui[(nN*nSim*1+1):(nN*nSim*2)],),
      "c:/simula/sim1_n100.res", ncolumns=14)
write(t(rEqui[(nN*nSim*2+1):(nN*nSim*3)],),
      "c:/simula/sim1_n50.res", ncolumns=14)
write(t(rEqui[(nN*nSim*3+1):(nN*nSim*4)],),
      "c:/simula/sim1_n20.res", ncolumns=14)

# Resumen estadístico de la simulación
archivos <- c( "c:/simula/sim1_n20.res",
               "c:/simula/sim1_n50.res",
               "c:/simula/sim1_n100.res",
               "c:/simula/sim1_n500.res" )

siml   <- list(n20=0, n50=0, n100=0, n500=0)
medias <- list(n20=0, n50=0, n100=0, n500=0)
ci.inf <- list(n20=0, n50=0, n100=0, n500=0)
ci.sup <- list(n20=0, n50=0, n100=0, n500=0)

plot.new()
for ( i in 1:length(archivos) )
{
  title(i,col.main="blue")

  rEqui <- scan( archivos[i] )
  dim(rEqui) <- c( 20, 20000 )
}

```

```

rEqui <- t( rEqui )

medias[[i]] <- aggregate(rEqui, list(n=rEqui[,1], mu=rEqui[,2]),
                        mean.na.rm)

ci.inf[[i]] <- aggregate(rEqui, list(n=rEqui[,1],mu=rEqui[,2]),
                        q2.5)
ci.sup[[i]] <- aggregate(rEqui, list(n=rEqui[,1],mu=rEqui[,2]),
                        q97.5)

siml[[i]] <- rEqui

title(i,col.main="white")
}

medias <- rbind(medias$n20[3:22],medias$n50[3:22],
               medias$n100[3:22],medias$n500[3:22])
ci.inf <- rbind(ci.inf$n20[3:22],ci.inf$n50[3:22],
               ci.inf$n100[3:22],ci.inf$n500[3:22])
ci.sup <- rbind(ci.sup$n20[3:22],ci.sup$n50[3:22],
               ci.sup$n100[3:22],ci.sup$n500[3:22])

names(medias) <- c( "n", "mu", "LR.val", "LR", "W.val", "W",
                  "LMR1.val", "LMR1", "LMR2.val", "LMR2",
                  "D.val", "D", "X2.val", "X2" )
names(ci.inf) <- names(medias)
names(ci.sup) <- names(medias)

tabla <- cbind( medias[,c("n","mu","LR","W","LMR1","LMR2","D","X2")],
               ci.inf[,c("LR.val","W.val","LMR1.val","LMR2.val",
                        "D.val","X2.val")],
               ci.sup[,c("LR.val","W.val","LMR1.val","LMR2.val",
                        "D.val","X2.val")] )

options(digits=4)

# Guardar resultados: Pruebas diagn3sticas de sobredispersi3n
write.table( tabla, file="c:/simula/siml_res.txt", quote=FALSE )

# Representaci3n gr3fica de las distribuciones muestrales de LR
plot.new() ; split.screen(c(2,2))
screen(1)
hist(rEqui[rEqui[,2]==0.3,3],
     main=expression(paste("Poisson(",lambda==0.3,")")),
     xlab="Raz3n de verisimilitud (LR)",ylab="Frecuencia")
screen(2)
hist(rEqui[rEqui[,2]==1,3],
     main=expression(paste("Poisson(",lambda==1,")")),
     xlab="Raz3n de verisimilitud (LR)",ylab="Frecuencia")
screen(3)
hist(rEqui[rEqui[,2]==5,3],
     main=expression(paste("Poisson(",lambda==5,")")),
     xlab="Raz3n de verisimilitud (LR)",ylab="Frecuencia")
screen(4)
hist(rEqui[rEqui[,2]==10,3],
     main=expression(paste("Poisson(",lambda==10,")")),
     xlab="Raz3n de verisimilitud (LR)",ylab="Frecuencia")

```

```
# Representación gráfica de las distribuciones muestrales de LR
plot.new() ; split.screen(c(2,2))
screen(1)
hist(rEqui[rEqui[,2]==0.3,5],
     main=expression(paste("Poisson(",lambda==0.3,"")),
     xlab="Prueba de Wald (W)",ylab="Frecuencia")

screen(2)
hist(rEqui[rEqui[,2]==1,5],
     main=expression(paste("Poisson(",lambda==1,"")),
     xlab="Prueba de Wald (W)",ylab="Frecuencia")
screen(3)
hist(rEqui[rEqui[,2]==5,5],
     main=expression(paste("Poisson(",lambda==5,"")),
     xlab="Prueba de Wald (W)",ylab="Frecuencia")
screen(4)
hist(rEqui[rEqui[,2]==10,3],
     main=expression(paste("Poisson(",lambda==10,"")),
     xlab="Prueba de Wald (W)",ylab="Frecuencia")

# Guardar entorno R para futuras sesiones
options(digits=7)
save.image("C:/Simula/sim1.RData")
```

Listado 2. Programa R para la Simulación 2.

```

# ===== #
# SIMULACIÓN 2: Muestras Negbin II #
# #
# Estudio comparativo de la potencia de los tests #
# diagnósticos de sobredispersión de uso habitual #
# ===== #

# Carga de funciones para el ajuste de modelos y
# tests de sobredispersión.
source("c:/simula/modelos.r")
source("c:/simula/tests.r")
library( MASS )

# Inicialización de parámetros de la simulación
set.seed(30)
tpo <- Sys.time()
plot.new()
n <- c( 500, 100, 50, 20 )
nN <- length( n )
mu <- c( 0.3, 1, 5 )
nB <- length(mu)
a <- rbind( c(1.2, 0.3, 0.15), c(4, 1, 0.5), c(20, 5, 2.5) )
nA <- ncol( a )
nSim <- 5000
rBN2 <- matrix(NA,nN*nB*nA*nSim,30)
rcto <- 0

# Simulación
for (h in 1:nN)
{
  # Generación de la variable independiente X
  x <- seq( from=-3, to=3, length=n[h] )

  for (i in 1:nB)
  {
    for (j in 1:nA)
    {
      for (k in 1:nSim)
      {

        # Contador de muestras generadas
        rcto <- rcto + 1
        title(rcto,col.main="blue")

        # Generación de la variable de respuesta Y bajo
        # distribución Negbin II
        y <- rnbinom( n[h], mu=mu[i], size=a[i,j] )

        # Ajuste de modelos para recuentos
        # (MRP, Negbin II y MRQP)
        poi <- glm.poisson( y, x )
        spoi <- summary(poi$modelo)
        nb <- glm.negbin( y, x )
        snb <- summary(nb$modelo)
        qpoi <- glm.quasipoisson( y, x )
        sqpoi <- summary(qpoi$modelo)
      }
    }
  }
}

```

```

# Tests de sobredispersión
o.lr <- LR( poi$LL, nb$LL )
o.w <- W( nb$modelo$theta, nb$modelo$SE.theta )
o.lm <- LM( y, poi$modelo$fit )
o.lmr <- LMR( y, poi$modelo$fit )

# Almacena vector de resultados
rBN2[rcto,1] <- n[h]
rBN2[rcto,2] <- mu[i]
rBN2[rcto,3] <- a[i,j]
rBN2[rcto,4] <- o.lr$val
rBN2[rcto,5] <- o.lr$sig
rBN2[rcto,6] <- o.w$val
rBN2[rcto,7] <- o.w$sig
rBN2[rcto,8] <- o.lmr$val[1]
rBN2[rcto,9] <- o.lmr$sig[1]
rBN2[rcto,10] <- o.lmr$val[2]
rBN2[rcto,11] <- o.lmr$sig[2]
rBN2[rcto,12] <- poi$modelo$dev
rBN2[rcto,13] <- (1 - pchisq( poi$modelo$dev,
                             poi$modelo$df.res )) <= 0.05
rBN2[rcto,14] <- poi$X2
rBN2[rcto,15] <- (1 - pchisq( poi$X2,
                             poi$modelo$df.res )) <= 0.05

rBN2[rcto,16] <- nb$modelo$theta
rBN2[rcto,17] <- nb$modelo$SE.theta
rBN2[rcto,18] <- nb$modelo$coef[1] # b0
rBN2[rcto,19] <- nb$modelo$coef[2] # b1
rBN2[rcto,20] <- snb$coef[1,2] # SE b0
rBN2[rcto,21] <- snb$coef[2,2] # SE b1
rBN2[rcto,22] <- poi$modelo$coef[1] # b0
rBN2[rcto,23] <- poi$modelo$coef[2] # b1
rBN2[rcto,24] <- spoi$coef[1,2] # SE b0
rBN2[rcto,25] <- spoi$coef[2,2] # SE b1
rBN2[rcto,26] <- qpoi$modelo$coef[1] # b0
rBN2[rcto,27] <- qpoi$modelo$coef[2] # b1
rBN2[rcto,28] <- sqpoi$coef[1,2] # SE b0
rBN2[rcto,29] <- sqpoi$coef[2,2] # SE b1
rBN2[rcto,30] <- sqpoi$dispersion

# Borra número de iteración
title(rcto,col.main="white")
}
}
}

print( TotTpo[1] <- difftime(Sys.time(),tpo) )

# Almacena en disco la matriz de resultados
write(t(rBN2[1:(nN*nA*nSim*1)],),
      "c:/simula/sim2_n500.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*1+1):(nN*nA*nSim*2)],),
      "c:/simula/sim2_n100.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*2+1):(nN*nA*nSim*3)],),
      "c:/simula/sim2_n50.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*3+1):(nN*nA*nSim*4)],),
      "c:/simula/sim2_n20.res",ncolumns=30)

```


ANEXOS

```

# Resumen estadístico de la simulación
archivos <- c( "c:/simula/sim2_n20.res",
               "c:/simula/sim2_n50.res",
               "c:/simula/sim2_n100.res",
               "c:/simula/sim2_n500.res")

sim2 <- list(n20=0, n50=0, n100=0, n500=0)
medias <- list(n20=0, n50=0, n100=0, n500=0)
desvia <- list(n20=0, n50=0, n100=0, n500=0)
ci.inf <- list(n20=0, n50=0, n100=0, n500=0)
ci.sup <- list(n20=0, n50=0, n100=0, n500=0)

plot.new()
for ( i in 1:length(archivos) )
{
  title(i,col.main="blue")

  rBN2 <- scan( archivos[i] )
  dim(rBN2) <- c( 40, 9000 )
  rBN2 <- t(rBN2)

  sim2[[i]] <- rBN2

  medias[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                   shape=rBN2[,3]),mean.na.rm)
  desvia[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                   shape=rBN2[,3]),sd.na.rm)
  ci.inf[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                   shape=rBN2[,3]),q2.5)
  ci.sup[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                   shape=rBN2[,3]),q97.5)

  title(i,col.main="white")
}

medias <- rbind(medias$n20[c(1:3,7:33)],medias$n50[c(1:3,7:33)],
               medias$n100[c(1:3,7:33)],medias$n500[c(1:3,7:33)])
desvia <- rbind(desvia$n20[c(1:3,7:33)],desvia$n50[c(1:3,7:33)],
               desvia$n100[c(1:3,7:33)],desvia$n500[c(1:3,7:33)])
ci.inf <- rbind(ci.inf$n20[c(1:3,7:33)],ci.inf$n50[c(1:3,7:33)],
               ci.inf$n100[c(1:3,7:33)],ci.inf$n500[c(1:3,7:33)])
ci.sup <- rbind(ci.sup$n20[c(1:3,7:33)],ci.sup$n50[c(1:3,7:33)],
               ci.sup$n100[c(1:3,7:33)],ci.sup$n500[c(1:3,7:33)])

names(medias) <- c( "n", "mu", "shape", "LR.val", "LR", "W.val", "W",
                  "LMR1.val", "LMR1", "LMR2.val", "LMR2", "D.val",
                  "D", "X2.val", "X2", "theta", "SE.theta",
                  "nb.b0", "nb.b1", "nb.se.b0", "nb.se.b1",
                  "poi.b0", "poi.b1", "poi.se.b0", "poi.se.b1",
                  "qpoi.b0", "qpoi.b1", "qpoi.se.b0", "qpoi.se.b1",
                  "qpoi.disp")
names(desvia) <- names(medias)
names(ci.inf) <- names(medias)
names(ci.sup) <- names(medias)

options(digits=4)

```

```

# Resultados: Pruebas diagnósticas de sobredispersión
tabla1 <- cbind( medias[, c("n","mu","shape","LR","W",
                          "LMR1","LMR2","D","X2")],
               ci.inf[, c("LR.val","W.val",
                          "LMR1.val","LMR2.val","D.val","X2.val")],
               ci.sup[, c("LR.val","W.val",
                          "LMR1.val","LMR2.val","D.val","X2.val")])
write.table( tabla1, file="c:/simula/sim2_res1.txt", quote=FALSE )

# Resultados: Estimación del parámetro de dispersión
tabla2 <- cbind( medias[, c("n","mu","shape","theta","qpoi.disp")],
               ci.inf[, c("theta","qpoi.disp")],
               ci.sup[, c("theta","qpoi.disp")] )
write.table( tabla2, file="c:/simula/sim2_res2.txt", quote=FALSE )

# Resultados: Estimación de los coeficientes del modelo de regresión
tabla3 <- cbind( medias[, c("n","mu","shape","nb.b0","poi.b0",
                          "qpoi.b0", "nb.b1","poi.b1","qpoi.b1")],
               ci.inf[, c("nb.b0","poi.b0",
                          "qpoi.b0","nb.b1","poi.b1","qpoi.b1")],
               ci.sup[, c("nb.b0","poi.b0",
                          "qpoi.b0","nb.b1","poi.b1","qpoi.b1")] )
write.table( tabla3, file="c:/simula/sim2_res3.txt", quote=FALSE )

# Resultados: Estimación del error estándar de los coeficientes
# del modelo de regresión
tabla4 <- cbind( medias[, c("n","mu","shape","nb.se.b0","poi.se.b0",
                          "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.inf[, c("nb.se.b0","poi.se.b0",
                          "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.sup[, c("nb.se.b0","poi.se.b0",
                          "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")] )
write.table( tabla4, file="c:/simula/sim2_res4.txt", quote=FALSE )

# Resultados: Estimación empírica del error estándar de los
# coeficientes a partir de la distribución muestral por Monte Carlo
tabla5 <- desvia[, c("n","mu","shape","nb.b0","poi.b0",
                   "qpoi.b0","nb.b1","poi.b1","qpoi.b1")]
write.table( tabla5, file="c:/simula/sim2_res5.txt", quote=FALSE )

# Guardar entorno R para futuras sesiones
options(digits=7)
save.image("C:/Simula/sim2.RData")

```

Listado 3. Programa R para la Simulación 3.

```

# ===== #
# SIMULACIÓN 3: Muestras por mezcla de dos Poisson #
# #
# Estudio comparativo de la potencia de los tests #
# diagnósticos de sobredispersión de uso habitual #
# ===== #

# Carga de funciones para el ajuste de modelos y
# tests de sobredispersión.
source("c:/simula/modelos.r")
source("c:/simula/tests.r")
library( MASS )

# Inicialización de parámetros de la simulación
set.seed(30)
tpo <- Sys.time()
plot.new()
n <- c( 500, 100, 50, 20 )
nN <- length( n )
mu <- rbind( c(0.2, 1), c(0.2, 3), c(3.2, 6) )
nB <- nrow(mu)
nSim <- 5000
r2Poi <- matrix(NA,nN*nB*nSim,30)
rcto <- 0

# Simulación
for (i in 1:nN)
{
  # Generación de la variable independiente X
  x <- seq( from=-3, to=3, length=n[i] )

  for (j in 1:nB)
  {
    for (k in 1:nSim)
    {
      # Contador de muestras generadas
      rcto <- rcto + 1
      title(rcto,col.main="blue")

      # Generación de la variable de respuesta Y bajo
      # dos distribuciones de Poisson con lambdas distintas
      y <- c( rpois( n[i]/2,lambda=mu[j,1] ),
              rpois( n[i]/2,lambda=mu[j,2] ) )

      # Ajuste de modelos para recuentos
      # (MRP, Negbin II y MRQP)
      poi <- glm.poisson( y, x )
      spoi <- summary(poi$modelo)
      nb <- glm.negbin( y, x )
      snb <- summary(nb$modelo)
      qpoi <- glm.quasipoisson( y, x )
      sqpoi <- summary(qpoi$modelo)

      # Tests de sobredispersión
      o.lr <- LR( poi$LL, nb$LL )
      o.w <- W( nb$modelo$theta, nb$modelo$SE.theta )
      o.lm <- LM( y, poi$modelo$fit )
      o.lmr <- LMR( y, poi$modelo$fit )
    }
  }
}

```

```

# Almacena vector de resultados
r2Poi[rcto,1] <- n[h]
r2Poi[rcto,2] <- mu[i]
r2Poi[rcto,3] <- a[i,j]
r2Poi[rcto,4] <- o.lr$val
r2Poi[rcto,5] <- o.lr$sig
r2Poi[rcto,6] <- o.w$val
r2Poi[rcto,7] <- o.w$sig
r2Poi[rcto,8] <- o.lmr$val[1]
r2Poi[rcto,9] <- o.lmr$sig[1]
r2Poi[rcto,10] <- o.lmr$val[2]
r2Poi[rcto,11] <- o.lmr$sig[2]
r2Poi[rcto,12] <- poi$modelo$dev
r2Poi[rcto,13] <- (1 - pchisq( poi$modelo$dev,
                             poi$modelo$df.res )) <= 0.05
r2Poi[rcto,14] <- poi$X2
r2Poi[rcto,15] <- (1 - pchisq( poi$X2,
                             poi$modelo$df.res )) <= 0.05

r2Poi[rcto,16] <- nb$modelo$theta
r2Poi[rcto,17] <- nb$modelo$SE.theta
r2Poi[rcto,18] <- nb$modelo$coef[1]      # b0
r2Poi[rcto,19] <- nb$modelo$coef[2]     # b1
r2Poi[rcto,20] <- snb$coef[1,2]        # SE b0
r2Poi[rcto,21] <- snb$coef[2,2]        # SE b1
r2Poi[rcto,22] <- poi$modelo$coef[1]    # b0
r2Poi[rcto,23] <- poi$modelo$coef[2]    # b1
r2Poi[rcto,24] <- spoi$coef[1,2]       # SE b0
r2Poi[rcto,25] <- spoi$coef[2,2]       # SE b1
r2Poi[rcto,26] <- qpoi$modelo$coef[1]   # b0
r2Poi[rcto,27] <- qpoi$modelo$coef[2]   # b1
r2Poi[rcto,28] <- sqpoi$coef[1,2]      # SE b0
r2Poi[rcto,29] <- sqpoi$coef[2,2]      # SE b1
r2Poi[rcto,30] <- sqpoi$dispersion

# Borra número de iteración
title(rcto,col.main="white")

}
}
}

print( TotTpo[1] <- difftime(Sys.time(),tpo) )

# Almacena en disco la matriz de resultados
write(t(r2Poi[1:(nB*nSim*1)],),
      "c:/simula/sim3_n500.res",ncolumns=30)
write(t(r2Poi[(nB*nSim*1+1):(nB*nSim*2)],),
      "c:/simula/sim3_n100.res",ncolumns=30)
write(t(r2Poi[(nB*nSim*2+1):(nB*nSim*3)],),
      "c:/simula/sim3_n50.res",ncolumns=30)
write(t(r2Poi[(nB*nSim*3+1):(nB*nSim*4)],),
      "c:/simula/sim3_n20.res",ncolumns=30)

# Resumen estadístico de la simulación
archivos <- c( "c:/simula/sim3_n20.res",
               "c:/simula/sim3_n50.res",
               "c:/simula/sim3_n100.res",
               "c:/simula/sim3_n500.res")

```

```

sim3 <- list(n20=0, n50=0, n100=0, n500=0)
medias <- list(n20=0, n50=0, n100=0, n500=0)
desvia <- list(n20=0, n50=0, n100=0, n500=0)
ci.inf <- list(n20=0, n50=0, n100=0, n500=0)
ci.sup <- list(n20=0, n50=0, n100=0, n500=0)

plot.new()
for ( i in 1:length(archivos) )
{
  title(i,col.main="blue")

  r2Poi <- scan( archivos[i] )
  dim(r2Poi) <- c( 40, 9000 )
  r2Poi <- t(r2Poi)

  sim3[[i]] <- r2Poi

  medias[[i]] <- aggregate(r2Poi,list(n=r2Poi[,1],mu=r2Poi[,2],
    shape=r2Poi[,3]),mean.na.rm)
  desvia[[i]] <- aggregate(r2Poi,list(n=r2Poi[,1],mu=r2Poi[,2],
    shape=r2Poi[,3]),sd.na.rm)
  ci.inf[[i]] <- aggregate(r2Poi,list(n=r2Poi[,1],mu=r2Poi[,2],
    shape=r2Poi[,3]),q2.5)
  ci.sup[[i]] <- aggregate(r2Poi,list(n=r2Poi[,1],mu=r2Poi[,2],
    shape=r2Poi[,3]),q97.5)

  title(i,col.main="white")
}

medias <- rbind(medias$n20[c(1:3,7:33)],medias$n50[c(1:3,7:33)],
  medias$n100[c(1:3,7:33)],medias$n500[c(1:3,7:33)])
desvia <- rbind(desvia$n20[c(1:3,7:33)],desvia$n50[c(1:3,7:33)],
  desvia$n100[c(1:3,7:33)],desvia$n500[c(1:3,7:33)])
ci.inf <- rbind(ci.inf$n20[c(1:3,7:33)],ci.inf$n50[c(1:3,7:33)],
  ci.inf$n100[c(1:3,7:33)],ci.inf$n500[c(1:3,7:33)])
ci.sup <- rbind(ci.sup$n20[c(1:3,7:33)],ci.sup$n50[c(1:3,7:33)],
  ci.sup$n100[c(1:3,7:33)],ci.sup$n500[c(1:3,7:33)])

names(medias) <- c( "n", "mu", "shape", "LR.val", "LR", "W.val", "W",
  "LMR1.val", "LMR1", "LMR2.val", "LMR2", "D.val",
  "D", "X2.val", "X2", "theta", "SE.theta",
  "nb.b0", "nb.b1", "nb.se.b0", "nb.se.b1",
  "poi.b0", "poi.b1", "poi.se.b0", "poi.se.b1",
  "qpoi.b0", "qpoi.b1", "qpoi.se.b0", "qpoi.se.b1",
  "qpoi.disp")
names(desvia) <- names(medias)
names(ci.inf) <- names(medias)
names(ci.sup) <- names(medias)

options(digits=4)

# Resultados: Pruebas diagn3sticas de sobredispersi3n
tablal <- cbind( medias[, c("n","mu","shape","LR","W",
  "LMR1","LMR2","D","X2")],
  ci.inf[, c("LR.val","W.val",
  "LMR1.val","LMR2.val","D.val","X2.val")],
  ci.sup[, c("LR.val","W.val",
  "LMR1.val","LMR2.val","D.val","X2.val")])
write.table( tablal, file="c:/simula/sim3_res1.txt", quote=FALSE )

```

```

# Resultados: Estimación del parámetro de dispersión
tabla2 <- cbind( medias[, c("n","mu","shape","theta","qpoi.disp")],
               ci.inf[, c("theta","qpoi.disp")],
               ci.sup[, c("theta","qpoi.disp")] )
write.table( tabla2, file="c:/simula/sim3_res2.txt", quote=FALSE )

# Resultados: Estimación de los coeficientes del modelo de regresión
tabla3 <- cbind( medias[, c("n","mu","shape","nb.b0","poi.b0",
                           "qpoi.b0", "nb.b1","poi.b1","qpoi.b1")],
               ci.inf[, c("nb.b0","poi.b0",
                           "qpoi.b0","nb.b1","poi.b1","qpoi.b1")],
               ci.sup[, c("nb.b0","poi.b0",
                           "qpoi.b0","nb.b1","poi.b1","qpoi.b1")] )
write.table( tabla3, file="c:/simula/sim3_res3.txt", quote=FALSE )

# Resultados: Estimación del error estándar de los coeficientes
# del modelo de regresión
tabla4 <- cbind( medias[, c("n","mu","shape","nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.inf[, c("nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.sup[, c("nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")] )
write.table( tabla4, file="c:/simula/sim3_res4.txt", quote=FALSE )

# Resultados: Estimación empírica del error estándar de los
# coeficientes a partir de la distribución muestral por Monte Carlo
tabla5 <- desvia[, c("n","mu","shape","nb.b0","poi.b0",
                   "qpoi.b0","nb.b1","poi.b1","qpoi.b1")]
write.table( tabla5, file="c:/simula/sim3_res5.txt", quote=FALSE )

# Guardar entorno R para futuras sesiones
options(digits=7)
save.image("C:/Simula/sim3.RData")

```

Listado 4. Programa R para la Simulación 4.

```

# ===== #
# SIMULACIÓN 4: Muestras con exceso de ceros #
# #
# Estudio comparativo de la potencia de los tests #
# diagnósticos de sobredispersión de uso habitual #
# ===== #

# Carga de funciones para el ajuste de modelos,
# tests de sobredispersión y generación de muestras
source("c:/simula/modelos.r")
source("c:/simula/tests.r")
source("c:/simula/muestreo.r")
library( MASS )

# Inicialización de parámetros de la simulación
set.seed(30)
tpo <- Sys.time()
plot.new()
n <- c( 500, 100, 50, 20 )
nN <- length( n )
mu <- rbind( c(0, -0.5), c(0, 0.5), c(0, 1.25) )
nB <- nrow(mu)
nSim <- 5000
rZIP <- matrix(NA,nN*nB*nSim,30)
rcto <- 0

# Simulación
for (i in 1:nN)
{
  # Generación de la variable independiente X
  x <- seq( from=-3, to=3, length=n[i] )

  for (j in 1:nB)
  {
    for (k in 1:nSim)
    {
      # Contador de muestras generadas
      rcto <- rcto + 1
      title(rcto,col.main="blue")

      # Generación de la variable de respuesta Y bajo
      # dos distribuciones de Poisson con lambdas distintas
      y <- gm.zip(n[i],b0=0,b1=1,
                 g0=mu[j,1],g1=mu[j,2],x)

      # Ajuste de modelos para recuentos
      # (MRP, Negbin II y MRQP)
      poi <- glm.poisson( y, x )
      spoi <- summary(poi$modelo)
      nb <- glm.negbin( y, x )
      snb <- summary(nb$modelo)
      qpoi <- glm.quasipoisson( y, x )
      sqpoi <- summary(qpoi$modelo)

      # Tests de sobredispersión
      o.lr <- LR( poi$LL, nb$LL )
      o.w <- W( nb$modelo$theta, nb$modelo$SE.theta )
      o.lm <- LM( y, poi$modelo$fit )
      o.lmr <- LMR( y, poi$modelo$fit )
    }
  }
}

```

```

# Almacena vector de resultados
rZIP[rcto,1] <- n[h]
rZIP[rcto,2] <- mu[i]
rZIP[rcto,3] <- a[i,j]
rZIP[rcto,4] <- o.lr$val
rZIP[rcto,5] <- o.lr$sig
rZIP[rcto,6] <- o.w$val
rZIP[rcto,7] <- o.w$sig
rZIP[rcto,8] <- o.lmr$val[1]
rZIP[rcto,9] <- o.lmr$sig[1]
rZIP[rcto,10] <- o.lmr$val[2]
rZIP[rcto,11] <- o.lmr$sig[2]
rZIP[rcto,12] <- poi$modelo$dev
rZIP[rcto,13] <- (1 - pchisq( poi$modelo$dev,
                             poi$modelo$df.res )) <= 0.05
rZIP[rcto,14] <- poi$X2
rZIP[rcto,15] <- (1 - pchisq( poi$X2,
                             poi$modelo$df.res )) <= 0.05

rZIP[rcto,16] <- nb$modelo$theta
rZIP[rcto,17] <- nb$modelo$SE.theta
rZIP[rcto,18] <- nb$modelo$coef[1]      # b0
rZIP[rcto,19] <- nb$modelo$coef[2]      # b1
rZIP[rcto,20] <- snb$coef[1,2]         # SE b0
rZIP[rcto,21] <- snb$coef[2,2]         # SE b1
rZIP[rcto,22] <- poi$modelo$coef[1]     # b0
rZIP[rcto,23] <- poi$modelo$coef[2]     # b1
rZIP[rcto,24] <- spoi$coef[1,2]        # SE b0
rZIP[rcto,25] <- spoi$coef[2,2]        # SE b1
rZIP[rcto,26] <- qpoi$modelo$coef[1]    # b0
rZIP[rcto,27] <- qpoi$modelo$coef[2]    # b1
rZIP[rcto,28] <- sqpoi$coef[1,2]       # SE b0
rZIP[rcto,29] <- sqpoi$coef[2,2]       # SE b1
rZIP[rcto,30] <- sqpoi$dispersion

# Borra número de iteración
title(rcto,col.main="white")

}
}
}

print( TotTpo[1] <- difftime(Sys.time(),tpo) )

# Almacena en disco la matriz de resultados
write(t(rZIP[1:(nB*nSim*1)],),
      "c:/simula/sim4_n500.res",ncolumns=30)
write(t(rZIP[(nB*nSim*1)+1:(nB*nSim*2)],),
      "c:/simula/sim4_n100.res",ncolumns=30)
write(t(rZIP[(nB*nSim*2)+1:(nB*nSim*3)],),
      "c:/simula/sim4_n50.res",ncolumns=30)
write(t(rZIP[(nB*nSim*3)+1:(nB*nSim*4)],),
      "c:/simula/sim4_n20.res",ncolumns=30)

# Resumen estadístico de la simulación
archivos <- c( "c:/simula/sim4_n20.res",
               "c:/simula/sim4_n50.res",
               "c:/simula/sim4_n100.res",
               "c:/simula/sim4_n500.res")

```


ANEXOS

```

sim4    <- list(n20=0, n50=0, n100=0, n500=0)
medias  <- list(n20=0, n50=0, n100=0, n500=0)
desvia  <- list(n20=0, n50=0, n100=0, n500=0)
ci.inf  <- list(n20=0, n50=0, n100=0, n500=0)
ci.sup  <- list(n20=0, n50=0, n100=0, n500=0)

plot.new()
for ( i in 1:length(archivos) )
{
  title(i,col.main="blue")

  rZIP <- scan( archivos[i] )
  dim(rZIP) <- c( 40, 9000 )
  rZIP <- t(rZIP)

  sim4[[i]] <- rZIP

  medias[[i]] <- aggregate(rZIP,list(n=rZIP[,1],mu=rZIP[,2],
    shape=rZIP[,3]),mean.na.rm)
  desvia[[i]] <- aggregate(rZIP,list(n=rZIP[,1],mu=rZIP[,2],
    shape=rZIP[,3]),sd.na.rm)
  ci.inf[[i]] <- aggregate(rZIP,list(n=rZIP[,1],mu=rZIP[,2],
    shape=rZIP[,3]),q2.5)
  ci.sup[[i]] <- aggregate(rZIP,list(n=rZIP[,1],mu=rZIP[,2],
    shape=rZIP[,3]),q97.5)

  title(i,col.main="white")
}

medias <- rbind(medias$n20[c(1:3,7:33)],medias$n50[c(1:3,7:33)],
  medias$n100[c(1:3,7:33)],medias$n500[c(1:3,7:33)])
desvia <- rbind(desvia$n20[c(1:3,7:33)],desvia$n50[c(1:3,7:33)],
  desvia$n100[c(1:3,7:33)],desvia$n500[c(1:3,7:33)])
ci.inf <- rbind(ci.inf$n20[c(1:3,7:33)],ci.inf$n50[c(1:3,7:33)],
  ci.inf$n100[c(1:3,7:33)],ci.inf$n500[c(1:3,7:33)])
ci.sup <- rbind(ci.sup$n20[c(1:3,7:33)],ci.sup$n50[c(1:3,7:33)],
  ci.sup$n100[c(1:3,7:33)],ci.sup$n500[c(1:3,7:33)])

names(medias) <- c( "n", "mu", "shape", "LR.val", "LR", "W.val", "W",
  "LMR1.val", "LMR1", "LMR2.val", "LMR2", "D.val",
  "D", "X2.val", "X2", "theta", "SE.theta",
  "nb.b0", "nb.b1", "nb.se.b0", "nb.se.b1",
  "poi.b0", "poi.b1", "poi.se.b0", "poi.se.b1",
  "qpoi.b0", "qpoi.b1", "qpoi.se.b0", "qpoi.se.b1",
  "qpoi.disp")
names(desvia) <- names(medias)
names(ci.inf) <- names(medias)
names(ci.sup) <- names(medias)

options(digits=4)

# Resultados: Pruebas diagn3sticas de sobredispersi3n
tablal <- cbind( medias[, c("n","mu","shape","LR","W",
  "LMR1","LMR2","D","X2")],
  ci.inf[, c("LR.val","W.val",
  "LMR1.val","LMR2.val","D.val","X2.val")],
  ci.sup[, c("LR.val","W.val",
  "LMR1.val","LMR2.val","D.val","X2.val")])
write.table( tablal, file="c:/simula/sim4_res1.txt", quote=FALSE )

```

```

# Resultados: Estimación del parámetro de dispersión
tabla2 <- cbind( medias[, c("n","mu","shape","theta","qpoi.disp")],
               ci.inf[, c("theta","qpoi.disp")],
               ci.sup[, c("theta","qpoi.disp")] )
write.table( tabla2, file="c:/simula/sim4_res2.txt", quote=FALSE )

# Resultados: Estimación de los coeficientes del modelo de regresión
tabla3 <- cbind( medias[, c("n","mu","shape","nb.b0","poi.b0",
                           "qpoi.b0", "nb.b1","poi.b1","qpoi.b1")],
               ci.inf[, c("nb.b0","poi.b0",
                           "qpoi.b0","nb.b1","poi.b1","qpoi.b1")],
               ci.sup[, c("nb.b0","poi.b0",
                           "qpoi.b0","nb.b1","poi.b1","qpoi.b1")] )
write.table( tabla3, file="c:/simula/sim4_res3.txt", quote=FALSE )

# Resultados: Estimación del error estándar de los coeficientes
# del modelo de regresión
tabla4 <- cbind( medias[, c("n","mu","shape","nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.inf[, c("nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")],
               ci.sup[, c("nb.se.b0","poi.se.b0",
                           "qpoi.se.b0","nb.se.b1","poi.se.b1","qpoi.se.b1")] )
write.table( tabla4, file="c:/simula/sim4_res4.txt", quote=FALSE )

# Resultados: Estimación empírica del error estándar de los
# coeficientes a partir de la distribución muestral por Monte Carlo
tabla5 <- desvia[, c("n","mu","shape","nb.b0","poi.b0",
                   "qpoi.b0","nb.b1","poi.b1","qpoi.b1")]
write.table( tabla5, file="c:/simula/sim4_res5.txt", quote=FALSE )

# Guardar entorno R para futuras sesiones
options(digits=7)
save.image("C:/Simula/sim3.RData")

```

Listado 5. Programa R para la Simulación 5.

```

# ===== #
# SIMULACIÓN 5: Muestras Negbin II #
# #
# Comparación técnicas para la corrección del EE #
# estimado por MRP en presencia de sobredispersión #
# ===== #

# Carga de funciones para el ajuste de modelos,
# tests de sobredispersión y procedimientos de
# corrección del EE estimado por MRP
source("c:/simula/modelos.r")
source("c:/simula/tests.r")
source("c:/simula/correc_ee.r")
library( MASS )

# Inicialización de parámetros de la simulación
set.seed(30)
tpo <- Sys.time()
plot.new()
n <- c( 500, 100, 50, 20 )
nN <- length( n )
mu <- c( 0.3, 1, 5 )
nB <- length(mu)
a <- rbind( c(1.2, 0.3, 0.15), c(4, 1, 0.5), c(20, 5, 2.5) )
nA <- ncol( a )
nSim <- 5000
rBN2 <- matrix(NA,nN*nB*nA*nSim,31)
rcto <- 0

# Simulación
for (h in 1:nN)
{
  # Generación de la variable independiente X
  x <- seq( from=-3, to=3, length=n[h] )

  for (i in 1:nB)
  {
    for (j in 1:nA)
    {
      for (k in 1:nSim)
      {

        # Contador de muestras generadas
        rcto <- rcto + 1
        title(rcto,col.main="blue")

        # Generación de la variable de respuesta Y bajo
        # distribución Negbin II
        y <- rnbinom( n[h], mu=mu[i], size=a[i,j] )

        # Ajuste de modelos para recuentos
        # (MRP, Negbin II y MRQP)
        poi <- glm.poisson( y, x )
        spoi <- summary(poi$modelo)
        qpoi <- glm.quasipoisson( y, x )
        sqpoi <- summary(qpoi$modelo)
      }
    }
  }
}

```

```

# Correcciones de los EE de los coeficientes MRP
se.boot <- SE.boot( y, x, n[h], nboot=200 )
se.jack <- SE.jack( y, x, n[h] )

# Almacena vector de resultados
rBN2[rcto,1] <- n[h]
rBN2[rcto,2] <- mu[i]
rBN2[rcto,3] <- a[i,j]
rBN2[rcto,4] <- poi$modelo$dev
rBN2[rcto,5] <- poi$X2
rBN2[rcto,6] <- poi$modelo$df.res
rBN2[rcto,7] <- spoi$coef[1,1] # b0
rBN2[rcto,8] <- spoi$coef[2,1] # b1
rBN2[rcto,9] <- spoi$coef[1,2] # SE b0
rBN2[rcto,10] <- spoi$coef[2,2] # SE b1
rBN2[rcto,11] <- sqpoi$coef[1,1] # b0
rBN2[rcto,12] <- sqpoi$coef[2,1] # b1
rBN2[rcto,13] <- sqpoi$coef[1,2] # SE b0
rBN2[rcto,14] <- sqpoi$coef[2,2] # SE b1
rBN2[rcto,15] <- qpoi$modelo$dev
rBN2[rcto,16] <- qpoi$X2
rBN2[rcto,17] <- sqpoi$dispersion
rBN2[rcto,18] <- se.boot$b0
rBN2[rcto,19] <- se.boot$b1
rBN2[rcto,20] <- se.boot$sse.b0
rBN2[rcto,21] <- se.boot$sse.b1
rBN2[rcto,22] <- se.jack$b0
rBN2[rcto,23] <- se.jack$b1
rBN2[rcto,24] <- se.jack$sse.b0
rBN2[rcto,25] <- se.jack$sse.b1
rBN2[rcto,26] <- spoi$coef[1,2]
* sqrt( poi$modelo$dev/poi$modelo$df.res )
rBN2[rcto,27] <- spoi$coef[1,2]
* sqrt( poi$X2/poi$modelo$df.res )
rBN2[rcto,28] <- spoi$coef[1,2]
* sqrt( sqpoi$dispersion )
rBN2[rcto,29] <- spoi$coef[2,2]
* sqrt( poi$modelo$dev/poi$modelo$df.res )
rBN2[rcto,30] <- spoi$coef[2,2]
* sqrt( poi$X2/poi$modelo$df.res )
rBN2[rcto,31] <- spoi$coef[2,2]
* sqrt( sqpoi$dispersion )

# Borra número de iteración
title(rcto,col.main="white")

}
}
}
print( TotTpo[1] <- difftime(Sys.time(),tpo) )

# Almacena en disco la matriz de resultados
write(t(rBN2[1:(nN*nA*nSim*1)],),
      "c:/simula/sim5_n500.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*1+1):(nN*nA*nSim*2)],),
      "c:/simula/sim5_n100.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*2+1):(nN*nA*nSim*3)],),
      "c:/simula/sim5_n50.res",ncolumns=30)
write(t(rBN2[(nN*nA*nSim*3+1):(nN*nA*nSim*4)],),
      "c:/simula/sim5_n20.res",ncolumns=30)

```

```

# Resumen estadístico de la simulación
archivos <- c( "c:/simula/sim5_n20.res",
              "c:/simula/sim5_n50.res",
              "c:/simula/sim5_n100.res",
              "c:/simula/sim5_n500.res")

sim5    <- list(n20=0, n50=0, n100=0, n500=0)
medias  <- list(n20=0, n50=0, n100=0, n500=0)
desvia  <- list(n20=0, n50=0, n100=0, n500=0)
ci.inf  <- list(n20=0, n50=0, n100=0, n500=0)
ci.sup  <- list(n20=0, n50=0, n100=0, n500=0)

plot.new()
for ( i in 1:length(archivos) )
{
  title(i,col.main="blue")

  rBN2 <- scan( archivos[i] )
  dim(rBN2) <- c( 40, 9000 )
  rBN2 <- t(rBN2)

  sim2[[i]] <- rBN2

  medias[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                     shape=rBN2[,3]),mean.na.rm)
  desvia[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                     shape=rBN2[,3]),sd.na.rm)
  ci.inf[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                     shape=rBN2[,3]),q2.5)
  ci.sup[[i]] <- aggregate(rBN2,list(n=rBN2[,1],mu=rBN2[,2],
                                     shape=rBN2[,3]),q97.5)

  title(i,col.main="white")
}

medias <- rbind(medias$n20[c(1:3,7:34)],medias$n50[c(1:3,7:34)],
               medias$n100[c(1:3,7:34)],medias$n500[c(1:3,7:34)])
desvia <- rbind(desvia$n20[c(1:3,7:34)],desvia$n50[c(1:3,7:34)],
               desvia$n100[c(1:3,7:34)],desvia$n500[c(1:3,7:34)])
ci.inf <- rbind(ci.inf$n20[c(1:3,7:34)],ci.inf$n50[c(1:3,7:34)],
               ci.inf$n100[c(1:3,7:34)],ci.inf$n500[c(1:3,7:34)])
ci.sup <- rbind(ci.sup$n20[c(1:3,7:34)],ci.sup$n50[c(1:3,7:34)],
               ci.sup$n100[c(1:3,7:34)],ci.sup$n500[c(1:3,7:34)])

names(medias) <- c( "n", "mu", "shape", "D", "X2", "df",
                  "poi.b0", "poi.b1", "poi.se.b0", "poi.se.b1",
                  "qpoi.b0", "qpoi.b1", "qpoi.se.b0", "qpoi.se.b1",
                  "qpoi.D", "qpoi.X2", "qpoi.disp",
                  "boot.b0", "boot.b1", "boot.se.b0", "boot.se.b1",
                  "jack.b0", "jack.b1", "jack.se.b0", "jack.se.b1",
                  "D.se.b0", "X2.se.b0", "disp.se.b0",
                  "D.se.b1", "X2.se.b1", "disp.se.b1")

names(desvia) <- names(medias)
names(ci.inf) <- names(medias)
names(ci.sup) <- names(medias)

options(digits=4)

```

```

# Resultados: Correcciones de los EE del coeficiente b0
tabla1 <- cbind( medias[,c("n","mu","shape",
                          "poi.se.b0","qpoi.se.b0",
                          "jack.se.b0","boot.se.b0",
                          "D.se.b0","X2.se.b0","disp.se.b0")],
               ci.inf[,c("poi.se.b0","qpoi.se.b0",
                         "jack.se.b0","boot.se.b0",
                         "D.se.b0","X2.se.b0","disp.se.b0")],
               ci.sup[,c("poi.se.b0","qpoi.se.b0",
                         "jack.se.b0","boot.se.b0",
                         "D.se.b0","X2.se.b0","disp.se.b0")] )
write.table( tabla1, file="c:/simula/sim5_res1.txt", quote=FALSE )

# Resultados: Correcciones de los EE del coeficiente b1
tabla2 <- cbind( medias[,c("n","mu","shape",
                          "poi.se.b1","qpoi.se.b1",
                          "jack.se.b1","boot.se.b1",
                          "D.se.b1","X2.se.b1","disp.se.b1")],
               ci.inf[,c("poi.se.b1","qpoi.se.b1",
                         "jack.se.b1","boot.se.b1",
                         "D.se.b1","X2.se.b1","disp.se.b1")],
               ci.sup[,c("poi.se.b1","qpoi.se.b1",
                         "jack.se.b1","boot.se.b1",
                         "D.se.b1","X2.se.b1","disp.se.b1")] )
write.table( tabla2, file="c:/simula/sim5_res2.txt", quote=FALSE )

# Guardar entorno R para futuras sesiones
options(digits=7)
save.image("C:/Simula/sim5.RData")

```

Listado 6. Funciones para la ampliación de las salidas de las funciones R estándar para el ajuste de modelos de regresión para recuentos.

```

# ----- #
# Regresión de Poisson #
# ----- #

glm.poisson <- function( y, x )
{
  # Ajuste del modelo de regresión
  modelo <- glm( y ~ x, family=poisson )

  # LogLikelihood
  # LL <- sum(dpois(y,modelo$fit,log=TRUE) * modelo$prior.weights)
  LL <- -1 * ( modelo$aic - 2*length(modelo$coef) ) / 2

  # X2 de Pearson
  X2 <- sum( ( y-modelo$fit) / sqrt(modelo$fit) )^2 )

  # G2
  G2 <- ( 2*sum(y[y>0] * log(y[y>0]/modelo$fit[y>0])
           - (y[y>0]-modelo$fit[y>0]) ) )
         + ( 2*sum(modelo$fit[y==0]) )

  # Retorno
  list( modelo=modelo, LL=LL, X2=X2, G2=G2 )
}

# ----- #
# Regresión con estimación Quasi-Poisson #
# ----- #

glm.quasipoisson <- function( y, x )
{
  # Ajuste del modelo de regresión
  modelo <- glm( y ~ x, family=quasipoisson )

  # LogLikelihood
  # LL <- sum(dpois(y,modelo$fit,log=TRUE) * qpoi$prior.weights)
  LL <- -1 * ( modelo$aic - 2*length(modelo$coef) ) / 2

  # X2 de Pearson
  X2 <- sum( ( y-modelo$fit) / sqrt(modelo$fit) )^2 )

  # G2
  G2 <- ( 2*sum(y[y>0] * log(y[y>0]/modelo$fit[y>0])
           - (y[y>0]-modelo$fit[y>0]) ) )
         + ( 2*sum(modelo$fit[y==0]) )

  # Retorno
  list( modelo=modelo, LL=LL, X2=X2, G2=G2 )
}

```

```
# ----- #  
# Regresión Binomial Negativa (Negbin II) #  
# ----- #  
  
glm.negbin <- function( y, x )  
{  
  
  # Ajuste del modelo de regresión  
  modelo <- glm.nb( y ~ x )  
  
  # LogLikelihood  
  LL <- modelo$twologlik / 2  
  
  # X2 de Pearson  
  X2 <- sum( ( y-modelo$fit ) / sqrt(modelo$fit) )^2 )  
  
  # G2  
  G2 <- ( 2*sum(y[y>0] * log(y[y>0]/modelo$fit[y>0])  
          - (y[y>0]-modelo$fit[y>0]) ) )  
        + ( 2*sum(modelo$fit[y==0]) )  
  
  # Retorno  
  list( modelo=modelo, LL=LL, X2=X2, G2=G2 )  
}
```


Listado 7. Funciones R para los tests diagn3sticos de sobredispersi3n y otras funciones de uso general en las simulaciones.

```

# ----- #
# Log-likelihood Ratio test (LR) #
# ----- #

LR <- function(poisson.LL,negbin.LL,pcrit=0.05)
{
  LR.val <- -2 * ( poisson.LL - negbin.LL )
  LR.p <- 1 - pchisq( LR.val, 1 )
  LR.sig <- ifelse( LR.p <= pcrit*2, 1, 0 )

  # Retorno
  list( val=LR.val, p=LR.p, sig=LR.sig )
}

# ----- #
# Test de Wald (W) #
# ----- #

W <- function(theta,SE.theta,pcrit=0.05)
{
  W.val <- theta / SE.theta
  W.p <- 1 - pchisq( W.val, 1 )
  W.sig <- ifelse( W.p <= pcrit*2, 1, 0 )

  # Retorno
  list( val=W.val, p=W.p, sig=W.sig )
}

# ----- #
# Lagrange Multiplier Test (LM) o "score" test #
# ----- #

LM <- function(y,poi.fit,pcrit=0.05)
{
  LM.val <- (sum((y-poi.fit)^2-y)/2) / sqrt(sum(1/(2*poi.fit^2)))
  LM.p <- 1 - pnorm( LM.val, mean=0, sd=1)
  LM.sig <- ifelse( LM.p <= pcrit, 1, 0 )

  # Retorno
  list( val=LM.val, p=LM.p, sig=LM.sig )
}

# ----- #
# Test basado en regresi3n OLS #
# (equivalente asint3tico al LM) #
# ----- #

LMR <- function(y,poi.fit,pcrit=0.05)
{
  z <- ((y-poi.fit)^2-y)/(sqrt(2)*poi.fit)

  # Binomial Negativa 1 (NB1)
  nb1.ols <- lm(z~+1)
  nb1.ols.val <- summary(nb1.ols)$coef[1,"t value"]
  nb1.ols.p <- summary(nb1.ols)$coef[1,"Pr(>|t|)"]
}

```

```

nb1.ols.sig <- ifelse( nb1.ols.p <= pcrit, 1, 0 )

# Binomial Negativa 2 (NB2)
nb2.ols <- lm(z~poi.fit-1)
nb2.ols.val <- summary(nb2.ols)$coef[1,"t value"]
nb2.ols.p <- summary(nb2.ols)$coef[1,"Pr(>|t|)"]
nb2.ols.sig <- ifelse( nb2.ols.p <= pcrit, 1, 0 )

# Retorno
list( nb1.ols=nb1.ols, nb2.ols=nb2.ols,
      val=c(nb1.ols.val,nb2.ols.val), p=c(nb1.ols.p,nb2.ols.p),
      sig=c(nb1.ols.sig,nb2.ols.sig) )
}

# ----- #
# Funciones generales para obtención de descriptivos con Aggregate #
# ----- #
mean.na.rm <- function(x) mean(x,na.rm=TRUE)
sd.na.rm <- function(x) sd(x,na.rm=TRUE)
q50 <- function(x) quantile(x,probs=c(0.50),na.rm=TRUE)
q2.5 <- function(x) quantile(x,probs=c(0.25),na.rm=TRUE)
q97.5 <- function(x) quantile(x,probs=c(0.975),na.rm=TRUE)

# ----- #
# Estimaciones directas del coeficiente de dispersión alfa #
# ----- #

overdisp <- function(glm)
{
  rr2 <- residuals(glm, type="response")^2
  rp2 <- residuals(glm, type="pearson")^2
  pw <- glm$prior.weights
  if (is.null(pw)) pw <- 1
  vi <- rp2/(pw*rr2) # 1/(mu + tau*mu^2)
  mu <- glm$fitted.values
  f <- glm$df.residual/length(mu)
  sum(rp2 - f*vi*mu)/(f*sum(vi*mu^2))
}

over.nb <- function(glm)
{
  rp2 <- residuals(glm, type="pearson")^2
  w <- glm$prior.weights
  if (is.null(w)) w <- 1
  mu <- glm$fitted.values
  h <- lm.influence(glm)$hat # diagonal elements of hat matrix
  sum(rp2 - (1-h)*w)/sum((1-h)*w*mu)
}

```

Listado 8. Funciones R para la corrección de los errores estándar estimados por Poisson en presencia de sobredispersión.

```

# ----- #
# Estimación bootstrap (no paramétrico) #
# de los coeficientes y sus EE #
# ----- #

SE.boot <- function( y, x, n, nboot=200 )
{
  b0.b <- double(nboot)
  b1.b <- double(nboot)
  for (i in 1:nboot)
  {
    k <- sample(1:n, replace = TRUE)
    xb <- x[k]
    yb <- y[k]
    coef <- glm(yb ~ xb, family=poisson)$coef

    b0.b[i] <- coef[1]
    b1.b[i] <- coef[2]
  }

  # Retorno
  list( b0=mean(b0.b), b1=mean(b1.b),
        se.b0=sd(b0.b), se.b1=sd(b1.b) )
}

# ----- #
# Estimación jackknife de los coeficientes y sus EE #
# ----- #

SE.jack <- function( y, x, n )
{
  b0.j <- double(n)
  b1.j <- double(n)
  for (i in 1:n)
  {
    xj <- x[-i]
    yj <- y[-i]
    coef <- glm(yj ~ xj, family=poisson)$coef

    b0.j[i] <- coef[1]
    b1.j[i] <- coef[2]
  }

  b0 <- mean(b0.j)
  b1 <- mean(b1.j)
  se.b0 <- sqrt( (n-1) * sum( (b0.j-b0)^2 ) /n )
  se.b1 <- sqrt( (n-1) * sum( (b1.j-b1)^2 ) /n )

  # Retorno
  list( b0=b0, b1=b1, se.b0=se.b0, se.b1=se.b1 )
}

```

Listado 9. Funciones R para el muestreo bajo diferentes distribuciones de recuentos utilizadas en las simulaciones.

```

# ===== #
# Funciones para generación de muestras bajo diferentes #
# configuraciones (distribuciones) de recuentos #
# ===== #

# ----- #
# Generación de muestras Poisson #
# ----- #

gm.poisson <- function( n=100, b0=0, b1=1, mu.x=0, sd.x=1 )
{
  x      <- rnorm(n,mu.x,sd.x)
  mu     <- exp( b0 + b1*x )
  y      <- rpois(n,mu)
  list( y=y, x=x )
}

# ----- #
# Generación de muestras Binomial Negativa (BN2) #
# ----- #

gm.bn <- function( n=100, b0=0, b1=1, alfa=0.5, mu.x=0, sd.x=1 )
{
  x      <- rnorm(n,mu.x,sd.x)
  mu     <- exp( b0 + b1*x )
  lambda <- rgamma(n,shape=1/alfa,scale=alfa*mu)
  y      <- rpois(n,lambda)
  list( y=y, x=x )
}

# ----- #
# Generación de muestras ZIP #
# ----- #

expit <- function(eta) exp(eta)/(1+exp(eta))

gm.zip <- function(n=100,b0=0,b1=1,g0=0,g1=1,x)
{
  pi     <- expit(b0+b1*x)
  mu     <- exp(g0+g1*x)
  y      <- numeric(n)
  z      <- (runif(n)<pi)
  y[z]   <- rpois(sum(z),mu[z])
  y[!z]  <- 0
  y
}

```

Referencias bibliográficas

- Aitchison, J. y Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4), 643-653.
- Aitkin, M., Anderson, D., Francis, B. y Hinde, J. (1989). *Statistical Modelling in GLIM*. New York: Oxford University Press.
- American Psychological Association (1994). *Publication manual of the American Psychological Association*. (4ª ed.). Washington, DC: Autor.
- American Psychological Association (2001). *Publication manual of the American Psychological Association*. (5ª ed.). Washington, DC: Autor.
- Anguera, M. T. (1989). Hacia una representación conceptual: teorías y modelos. En J. Mayor y J. L. Pinillos (Eds.), *Tratado de Psicología general. Vol I: Teoría, historia y método* (pp. 543-580). Madrid: Alhambra.
- Ato, M. y López, J. J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis.
- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. L. y Rodrigo, M. F. (2000a). *Del contraste de hipótesis al modelado estadístico*. Terrassa: CBS.
- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. L. y Rodrigo, M. F. (2000b). *Modelo lineal generalizado*. Terrassa: CBS.
- Bartholomew, D. J. (1995). What is statistics? *Journal of the Royal Statistical Society-Series A*, 158, 1-20.
- Bates, G. y Neyman, J. (1951). Contributions to the theory of accident proneness. II: True or false contagion. *University of California Publications in Statistics*, 215-253.
- Biggeri, A., Marchi, M., Lagazio, C., Martuzzi, M. y Böhning, D. (2000). Non-parametric maximum likelihood estimators for disease mapping [Versión electrónica]. *Statistics in Medicine*, 19(17-18), 2539-2554.
- Box, G. E. P. y Jenkins, G. M. (1976). *Time series analysis: Forecasting and Control*. San Francisco: Holden-Day.

- Brännäs, K. y Rosenqvist, G. (1994). Semiparametric estimation of heterogeneous count data models. *European Journal of Operational Research*, 76, 247-258.
- Breslow, N. (1990). Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models [Versión electrónica]. *Journal of the American Statistical Association*, 85(410), 565-571.
- Breslow, N. (1996). Generalized linear models: checking assumptions and strengthening conclusions. *Statistica Applicata*, 8, 23-41.
- Cameron, A. C. y Trivedi, P. K. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests [Versión electrónica]. *Journal of Applied Econometrics*, 1, 29-53.
- 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. (1998). Regression Analysis of Count Data. *Econometric Society Monographs*, 30. Cambridge: Cambridge University Press.
- Consul, P. C. (1989). *Generalized Poisson distributions*. New York: Marcel Dekker.
- Consul, P. C. y Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics - Theory and Methods*, 21, 89-109.
- Cortés, J. y Martínez, A. (1996). *Diccionario de filosofía en CD ROM*. Barcelona: Herder.
- Cox, J. L., Heyse, J. F. y Tukey, J. W. (2000). Efficacy Estimates from Parasite Count Data That Include Zero Counts [Versión electrónica]. *Experimental Parasitology*, 96(1), 1-8.
- Crepon, B. y Duguet, E. (1997). Research and development, competition and innovation, Pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity [Versión electrónica]. *Journal of Econometrics*, 79(2), 355-378.
- Czado, C. y Munk, A. (2000). *Noncanonical links in generalized linear models*. Recuperado el 5/4/2000 de <http://www.math.yorku.ca/who/faculty/czado/linkn3.ps>.
- Dean, C. B., Lawless, J. F. y Willmot, G. E. (1989). A mixed Poisson-inverse Gaussian regression model. *Canadian Journal of Statistics*, 17(2), 171-181.

- Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman & Hall.
- Fader, P. S. y Hardie, B. G. S. (2000). A note on modelling underreported Poisson counts [Versión electrónica]. *Journal of Applied Statistics*, 27(8), 953-964.
- Fahrmeir, L. y Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors [Versión electrónica]. *Applied Statistics*, 50(2), 201-220.
- Fahrmeir, L. y Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. (2ª ed.). New York: Springer-Verlag.
- Famoye, F. (1993). Restricted generalized Poisson regression. *Communications in Statistics - Theory and Methods*, 22, 1335-1354.
- García-Crespo, D. (2001). Promotions in the Spanish labour market: differences by gender [Versión electrónica]. *Oxford Bulletin of Economics and Statistics*, 63(5), 599-615.
- Gardner, W., Mulvey, E. y Shaw, E. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392-404.
- Gill, Jeff (2001). Generalized linear models: A unified approach. *Sage University Papers on Quantitative Applications in the Social Sciences*, 07-134. Thousand Oaks, CA: Sage.
- Goldstein, R. y Harrell, F. (1998). Survival analysis, software. En P. Armitage y T. Colton (Eds.), *Encyclopedia of Biostatistics* (Obtenido el 12/01/02 de <http://www.wiley.co.uk/wileychi/eob/sample6.pdf>). London: Wiley.
- Gourieroux, C. y Magnac, T. (1997). Duration, transition in count data models [Versión electrónica]. *Journal of Econometrics*, 79, 195-199.
- Gourieroux, C., Monfort, A. y Trognon, A. (1984a). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52, 701-720.
- Gourieroux, C., Monfort, A. y Trognon, A. (1984b). Pseudo-maximum likelihood methods: theory. *Econometrica*, 52, 681-700.
- Gourieroux, C. y Visser, M. (1997). A count data model with unobserved heterogeneity [Versión electrónica]. *Journal of Econometrics*, 79, 247-268.
- Greene, W. H. (2000). *Econometric analysis*. (4ª ed.). New York: Prentice Hall.

- Grogger, J. T. y Carson, R. T. (1991). Models for truncated counts. *Journal of Applied Econometrics*, 6, 225-238.
- Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. *Journal of Business and Economic Statistics*, 9(2), 215-222.
- Gurmu, S., Rilstone, P. y Stern, S. (1998). Semiparametric estimation of count regression models. *Journal of Econometrics*, 88(1), 123-150.
- Gurmu, S. y Trivedi, P. K. (1992). Overdispersion tests for truncated Poisson regression models. *Journal of Econometrics*, 54, 347-370.
- Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical bayes estimation [Versión electrónica]. *Accident Analysis and Prevention*, 33(6), 799-808.
- Hausman, J., Hall, B. H. y Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship [Versión electrónica]. *Econometrica*, 52(4), 909-938.
- Heo, G. (2000). *Log-Linear Regression for Poisson Counts*. Recuperado el 12/1/2002 de <http://www.stat.ualberta.ca/people/heo/lec22.pdf>.
- Hutcheson, G. D. y Sofroniou, N. (1999). *The Multivariate Social Scientist. Introductory Statistics Using Generalized Linear Models*. London: Sage.
- Jáñez, J. (1989). *Fundamentos de Psicología matemática*. Madrid: Ediciones Pirámide.
- Johnson, N. L., Kotz, S. y Balakrishnan, N. (1994). *Continuous univariate distributions. Vol. I. (2ª ed.)*. New York: John Wiley.
- Judd, C. M. y McClelland, G. H. (1989). *Data Analysis: A Model-Comparison Approach*. San Diego, CA: Harcourt Brace Jovanovich.
- King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for exponential Poisson regression model [Versión electrónica]. *American Journal of Political Science*, 32(3), 838-863.
- King, G. (1989a). A seemingly unrelated Poisson regression model [Versión electrónica]. *Sociological Methods and Research*, 17, 235-255.
- King, G. (1989b). Variance specification in event count models: From restrictive assumptions to a generalized estimator [Versión electrónica]. *American Journal of Political Science*, 33(3), 762-784.

- King, G. y Signorino, C. S. (1995). *The generalization in the generalized event count model, with comments on Achen, Amato, and Londregan [Versión electrónica]*. Recuperado el 22/11/2001 de <http://wizard.ucr.edu/polmeth/working-papers95/king95b.html>.
- Kleinbaum, D. G., Kupper, L. L. y Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariate Methods*. (2ª ed.). Belmont, CA: Duxbury Press.
- Krzanowski, W. J. (1998). *An Introduction to Statistical Modelling*. London: Arnold.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, MI: University of Chicago Press.
- Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15, 209-225.
- Lemeshow, S. y Hosmer, D. W. (1998). Logistic regression. En P. Armitage y T. Colton (Eds.), *Encyclopedia of Biostatistics* (Obtenido el 12/01/02 de <http://www.wiley.co.uk/wileychi/eob/sample6.pdf>). London: Wiley.
- Liao, T. F. (2000). Estimated precision for predictions from generalized linear models in sociological research. *Quality & Quantity*, 34, 137-152.
- Lichstein, K. L., Riedel, B. W., Wilson, N. M., Lester, K. W. y Aguilard, R. N. (2001). Relaxation and sleep compression for late-life insomnia: a placebo-controlled trial. *Journal of Consulting and Clinical Psychology*, 69(2), 227-239.
- Lindsey, J. K. (1995a). *Introductory Statistics: A Modelling Approach*. New York: Oxford University Press.
- Lindsey, J. K. (1995b). *Modelling frequency and count data*. Oxford: Clarendon Press.
- Lindsey, J. K. (1997). *Applying Generalized Linear Models*. New York: Springer-Verlag.
- Lindsey, J. K. (1998). Counts and times to events. *Statistics in Medicine*, 17(15-16), 1745-1751.

- Lindsey, J. K., Jones, B. y Jarvis, P. (2001). Some statistical issues in modelling pharmacokinetic data [Versión electrónica]. *Statistics in Medicine*, 20(17-18), 2775-2783.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Losilla, J. M. (1995). *Proyecto docente de la asignatura «Software» en Psicología*. (Manuscrito no publicado). Barcelona: Universitat Autònoma de Barcelona, Departament de Psicologia de la Salut.
- Lunneborg, C. E. (1994). *Modelling Experimental and Observational Data*. Belmont, CA: Duxbury Press.
- MacNaughton, K. L. y Rodrigue, J. R. (2001). Predicting adherence to recommendations by parents of clinic-referred. *Journal of Consulting and Clinical Psychology*, 69(2), 262-270.
- McCullagh, P. y Nelder, J. A. (1989). *Generalized linear models*. (2ª ed.). London: Chapman & Hall.
- Meliciani, V. (2000). The relationship between R&D, investment and patents: a panel data analysis [Versión electrónica]. *Applied Economics*, 32, 1429-1437.
- Melkersson, M. y Rooth, D.-O. (2000). Modelling female fertility using inflated count data models [Versión electrónica]. *Journal of Population Economics*, 13, 189-203.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341-365.
- Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models [Versión electrónica]. *Journal of Applied Econometrics*, 12, 337-350.
- Munnich, E., Landau, B. y Doshier, B. A. (2001). Spatial language spatial representation: a cross-linguistic comparison. *Cognition*, 81, 171-207.
- Nakashima, E. (1997). Some methods for estimation in a negative-binomial model [Versión electrónica]. *Annals of the Institute of Statistical Mathematics*, 49(1), 101-115.
- Navarro, A., Utzet, F., Puig, P., Caminal, J. y Martín, M. (2001). La distribución binomial negativa frente a la de Poisson en el análisis de fenómenos recurrentes [Versión electrónica]. *Gaceta Sanitaria*, 15(5), 447-452.

- Nelder, J. A. y Wedderburn, W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society-Series A*, 135, 370-384.
- Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates [Versión electrónica]. *Journal of Quantitative Criminology*, 16(1), 21-43.
- Palmer, J., Losilla, J. M., Sospedra, M. J., Sesé, A., Montaña, J., Jiménez, R. y Cajal, A. (2001, Septiembre). La sobredispersión en el MLG. En A. Palmer (Coord.), *Modelo Lineal Generalizado II*. Simposio realizado en el VII Congreso de Metodología de las Ciencias Sociales y de la Salud, Madrid.
- Poortema, K. (1999). On modelling overdispersion of counts. *Statistica Neerlandica*, 53(1), 5-20.
- Reese, R. A. (1986). Data analysis: The need for models? *The Statistician*, 35, 199-206.
- Ribeiro, F. F. (1999). Underreporting of purchases of port wine [Versión electrónica]. *Journal of Applied Statistics*, 26(4), 185-494.
- Rodrigo, M. F. (2000). *Proyecto Docente de la Asignatura Análisis de Datos*. (Manuscrito no publicado). Valencia: Universitat de València. Departament de Metodologia de les Ciències del Comportament.
- Rodríguez, G. (2002). *Lecture Notes on Generalized Linear Models*. Recuperado el 22/2/2002 de <http://data.princeton.edu/wws509/notes>.
- Rothenberg, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica*, 52, 827-842.
- Saha, A. y Dong, D. (1997). Estimating nested count data models [Versión electrónica]. *Oxford Bulletin of Economics and Statistics*, 59(3), 423-430.
- Scollnik, D. P. M. (1995). Bayesian analysis of two overdispersed Poisson models [Versión electrónica]. *Biometrics*, 51, 1117-1126.
- Sejnowski, T. J., Koch, C. y Churchland, P. S. (1990). Computational neuroscience. En S.J.Hanson y C. R. Olson (Eds.), *Connectionist Modelling and Brain Function: The Developing Interface* (pp. 5-35). Cambridge, MA: The MIT Press.
- Shankar, V., Milton, J. y Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry [Versión electrónica]. *Accident Analysis and Prevention*, 29(6), 829-837.

- Sturman, M. C. (1999). Multiple approaches to analyzing count data in studies of individual differences: The propensity for Type I errors, illustrated with the case of absenteeism prediction. *Educational and Psychological Measurement*, 59(3), 414-430.
- Suárez, C. P. (1996). Neurociencia y computación neuronal: una perspectiva. En C.P.Suárez y J. Regidor (Eds.), *Neurociencia y computación neuronal* (pp. 251-266). Las Palmas de Gran Canaria: Universidad de las Palmas de Gran Canaria. Sociedad Canaria de Neurociencia.
- Trivedi, P. K. (1997). Econometric models of event counts [Versión electrónica]. *Journal of Applied Econometrics*, 12, 199-201.
- Tu, W. y Zhou, X. (1999). A wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in Medicine*, 18, 2749-2761.
- van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, 51, 738-743.
- Vives, J. y Losilla, J. M. (2001, Septiembre). Las variables de recuento en la investigación en Psicología. En A.Palmer (Coord.), *Modelo Lineal Generalizado II*. Simposio realizado en el VII Congreso de Metodología de las Ciencias Sociales y de la Salud, Madrid.
- Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-334.
- Wang, W. y Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression [Versión electrónica]. *Journal of Population Economics*, 10, 273-283.
- Winkelmann, R. (1995). Duration dependence and dispersion in count data models. *Journal of Business and Economic Statistics*, 13(4), 467-474.
- Winkelmann, R. (2000). *Econometric Analysis of Count Data*. (3ª ed.). Berlin: Springer-Verlag.
- Winkelmann, R. y Zimmermann, K. F. (1991). A new approach for modeling economic count data. *Economic Letters*, 37, 139-143.
- Winkelmann, R. y Zimmermann, K. F. (1995). Recent developments in count data modelling: theory and application. *Journal of Economic Surveys*, 9(1), 1-24.

- Xekalaki, E. (1983). The univariate generalised Warning distribution in relation to accident theory: proneness, spells or contagion. *Biometrics*, 39, 887-895.
- Yau, K. K. W. y Lee, A. H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury programme [Versión electrónica]. *Statistics in Medicine*, 20, 2907-2920.
- Yen, S. T. (1999). Gaussian versus count-data hurdle models: cigarette consumption by women in the US [Versión electrónica]. *Applied Economic Letters*, 6, 73-76.

- Goldstein, R. y Harrell, F. (1998). Survival analysis, software. En P.Armitage y T. Colton (Eds.), *Encyclopedia of Biostatistics* (Obtenido el 12/01/02 de <http://www.wiley.co.uk/wileychi/eob/sample6.pdf>). London: Wiley.
- Lemeshow, S. y Hosmer, D. W. (1998). Logistic regression. En P.Armitage y T. Colton (Eds.), *Encyclopedia of Biostatistics* (Obtenido el 12/01/02 de <http://www.wiley.co.uk/wileychi/eob/sample6.pdf>). London: Wiley.
- Vives, J. y Losilla, J. M. (2001, Septiembre). Las variables de recuento en la investigación en Psicología. En A.Palmer (Coord.), *Modelo Lineal Generalizado II*. Simposio realizado en el VII Congreso de Metodología de las Ciencias Sociales y de la Salud, Madrid.
- Palmer, J., Losilla, J. M., Sospedra, M. J., Sesé, A., Montaña, J., Jiménez, R. y Cajal, A. (2001, Septiembre). La sobredispersión en el MLG. En A.Palmer (Coord.), *Modelo Lineal Generalizado II*. Simposio realizado en el VII Congreso de Metodología de las Ciencias Sociales y de la Salud, Madrid.
- Aitchison, J. y Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4), 643-653.
- Aitkin, M., Anderson, D., Francis, B. y Hinde, J. (1989). *Statistical Modelling in GLIM*. New York: Oxford University Press.
- American Psychological Association (1994). *Publication manual of the American Psychological Association*. (4ª ed.). Washington, DC: Autor.
- American Psychological Association (2001). *Publication manual of the American Psychological Association*. (5ª ed.). Washington, DC: Autor.
- Anguera, M. T. (1989). Hacia una representación conceptual: teorías y modelos. En J.Mayor y J. L. Pinillos (Eds.), *Tratado de Psicología general. Vol I: Teoría, historia y método* (pp. 543-580). Madrid: Alhambra.
- Ato, M. y López, J. J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis.
- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. L. y Rodrigo, M. F. (2000a). *Del contraste de hipótesis al modelado estadístico*. Terrassa: CBS.

- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. L. y Rodrigo, M. F. (2000b). *Modelo lineal generalizado*. Terrassa: CBS.
- Bartholomew, D. J. (1995). What is statistics? *Journal of the Royal Statistical Society-Series A*, 158, 1-20.
- Bates, G. y Neyman, J. (1951). Contributions to the theory of accident proneness. II: True or false contagion. *University of California Publications in Statistics*, 215-253.
- Biggeri, A., Marchi, M., Lagazio, C., Martuzzi, M. y Böhning, D. (2000). Non-parametric maximum likelihood estimators for disease mapping [Versión electrónica]. *Statistics in Medicine*, 19(17-18), 2539-2554.
- Box, G. E. P. y Jenkins, G. M. (1976). *Time series analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Brännäs, K. y Rosenqvist, G. (1994). Semiparametric estimation of heterogeneous count data models. *European Journal of Operational Research*, 76, 247-258.
- Breslow, N. (1990). Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models [Versión electrónica]. *Journal of the American Statistical Association*, 85(410), 565-571.
- Breslow, N. (1996). Generalized linear models: checking assumptions and strengthening conclusions. *Statistica Applicata*, 8, 23-41.
- Cameron, A. C. y Trivedi, P. K. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests [Versión electrónica]. *Journal of Applied Econometrics*, 1, 29-53.
- 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. (1998). *Regression Analysis of Count Data*. *Econometric Society Monographs*, 30. Cambridge: Cambridge University Press.
- Consul, P. C. (1989). *Generalized Poisson distributions*. New York: Marcel Dekker.
- Consul, P. C. y Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics - Theory and Methods*, 21, 89-109.

- Cortés, J. y Martínez, A. (1996). *Diccionario de filosofía en CD ROM*. Barcelona: Herder.
- Cox, J. L., Heyse, J. F. y Tukey, J. W. (2000). Efficacy Estimates from Parasite Count Data That Include Zero Counts [Versión electrónica]. *Experimental Parasitology*, 96(1), 1-8.
- Crepon, B. y Duguet, E. (1997). Research and development, competition and innovation, Pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity [Versión electrónica]. *Journal of Econometrics*, 79(2), 355-378.
- Czado, C. y Munk, A. (2000). *Noncanonical links in generalized linear models*. Recuperado el 5/4/2000 de <http://www.math.yorku.ca/who/faculty/czado/linkn3.ps>.
- Dean, C. B., Lawless, J. F. y Willmot, G. E. (1989). A mixed Poisson-inverse Gaussian regression model. *Canadian Journal of Statistics*, 17(2), 171-181.
- Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman & Hall.
- Fader, P. S. y Hardie, B. G. S. (2000). A note on modelling underreported Poisson counts [Versión electrónica]. *Journal of Applied Statistics*, 27(8), 953-964.
- Fahrmeir, L. y Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors [Versión electrónica]. *Applied Statistics*, 50(2), 201-220.
- Fahrmeir, L. y Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. (2ª ed.). New York: Springer-Verlag.
- Famoye, F. (1993). Restricted generalized Poisson regression. *Communications in Statistics - Theory and Methods*, 22, 1335-1354.
- García-Crespo, D. (2001). Promotions in the Spanish labour market: differences by gender [Versión electrónica]. *Oxford Bulletin of Economics and Statistics*, 63(5), 599-615.
- Gardner, W., Mulvey, E. y Shaw, E. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392-404.

- Gill, Jeff (2001). Generalized linear models: A unified approach. *Sage University Papers on Quantitative Applications in the Social Sciences*, 07-134. Thousand Oaks, CA: Sage.
- Gourieroux, C. y Magnac, T. (1997). Duration, transition in count data models [Versión electrónica]. *Journal of Econometrics*, 79, 195-199.
- Gourieroux, C., Monfort, A. y Trognon, A. (1984a). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52, 701-720.
- Gourieroux, C., Monfort, A. y Trognon, A. (1984b). Pseudo-maximum likelihood methods: theory. *Econometrica*, 52, 681-700.
- Gourieroux, C. y Visser, M. (1997). A count data model with unobserved heterogeneity [Versión electrónica]. *Journal of Econometrics*, 79, 247-268.
- Greene, W. H. (2000). *Econometric analysis*. (4ª ed.). New York: Prentice Hall.
- Grogger, J. T. y Carson, R. T. (1991). Models for truncated counts. *Journal of Applied Econometrics*, 6, 225-238.
- Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. *Journal of Business and Economic Statistics*, 9(2), 215-222.
- Gurmu, S., Rilstone, P. y Stern, S. (1998). Semiparametric estimation of count regression models. *Journal of Econometrics*, 88(1), 123-150.
- Gurmu, S. y Trivedi, P. K. (1992). Overdispersion tests for truncated Poisson regression models. *Journal of Econometrics*, 54, 347-370.
- Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical bayes estimation [Versión electrónica]. *Accident Analysis and Prevention*, 33(6), 799-808.
- Hausman, J., Hall, B. H. y Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship [Versión electrónica]. *Econometrica*, 52(4), 909-938.
- Heo, G. (2000). *Log-Linear Regression for Poisson Counts*. Recuperado el 12/1/2002 de <http://www.stat.ualberta.ca/people/heo/lec22.pdf>.
- Hutcheson, G. D. y Sofroniou, N. (1999). *The Multivariate Social Scientist. Introductory Statistics Using Generalized Linear Models*. London: Sage.
- Jáñez, J. (1989). *Fundamentos de Psicología matemática*. Madrid: Ediciones Pirámide.

- Johnson, N. L., Kotz, S. y Balakrishnan, N. (1994). *Continuous univariate distributions. Vol. I. (2ª ed.)*. New York: John Wiley.
- Judd, C. M. y McClelland, G. H. (1989). *Data Analysis: A Model-Comparison Approach*. San Diego, CA: Harcourt Brace Jovanovich.
- King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for exponential Poisson regression model [Versión electrónica]. *American Journal of Political Science*, 32(3), 838-863.
- King, G. (1989a). A seemingly unrelated Poisson regression model [Versión electrónica]. *Sociological Methods and Research*, 17, 235-255.
- King, G. (1989b). Variance specification in event count models: From restrictive assumptions to a generalized estimator [Versión electrónica]. *American Journal of Political Science*, 33(3), 762-784.
- King, G. y Signorino, C. S. (1995). *The generalization in the generalized event count model, with comments on Achen, Amato, and Londregan* [Versión electrónica]. Recuperado el 22/11/2001 de <http://wizard.ucr.edu/polmeth/working-papers95/king95b.html>.
- Kleinbaum, D. G., Kupper, L. L. y Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariate Methods. (2ª ed.)*. Belmont, CA: Duxbury Press.
- Krzanowski, W. J. (1998). *An Introduction to Statistical Modelling*. London: Arnold.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, MI: University of Chicago Press.
- Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15, 209-225.
- Liao, T. F. (2000). Estimated precision for predictions from generalized linear models in sociological research. *Quality & Quantity*, 34, 137-152.
- Lichstein, K. L., Riedel, B. W., Wilson, N. M., Lester, K. W. y Aguilard, R. N. (2001). Relaxation and sleep compression for late-life insomnia: a placebo-controlled trial. *Journal of Consulting and Clinical Psychology*, 69(2), 227-239.

- Lindsey, J. K. (1995a). *Introductory Statistics: A Modelling Approach*. New York: Oxford University Press.
- Lindsey, J. K. (1995b). *Modelling frequency and count data*. Oxford: Clarendon Press.
- Lindsey, J. K. (1997). *Applying Generalized Linear Models*. New York: Springer-Verlag.
- Lindsey, J. K. (1998). Counts and times to events. *Statistics in Medicine*, 17(15-16), 1745-1751.
- Lindsey, J. K., Jones, B. y Jarvis, P. (2001). Some statistical issues in modelling pharmacokinetic data [Versión electrónica]. *Statistics in Medicine*, 20(17-18), 2775-2783.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Losilla, J. M. (1995). *Proyecto docente de la asignatura «Software» en Psicología*. (Manuscrito no publicado). Barcelona: Universitat Autònoma de Barcelona, Departament de Psicologia de la Salut.
- Lunneborg, C. E. (1994). *Modelling Experimental and Observational Data*. Belmont, CA: Duxbury Press.
- MacNaughton, K. L. y Rodrigue, J. R. (2001). Predicting adherence to recommendations by parents of clinic-referred. *Journal of Consulting and Clinical Psychology*, 69(2), 262-270.
- McCullagh, P. y Nelder, J. A. (1989). *Generalized linear models*. (2ª ed.). London: Chapman & Hall.
- Meliciani, V. (2000). The relationship between R&D, investment and patents: a panel data analysis [Versión electrónica]. *Applied Economics*, 32, 1429-1437.
- Melkersson, M. y Rooth, D.-O. (2000). Modelling female fertility using inflated count data models [Versión electrónica]. *Journal of Population Economics*, 13, 189-203.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341-365.
- Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models [Versión electrónica]. *Journal of Applied Econometrics*, 12, 337-350.

- Munnich, E., Landau, B. y Doshier, B. A. (2001). Spatial language spatial representation: a cross-linguistic comparison. *Cognition*, 81, 171-207.
- Nakashima, E. (1997). Some methods for estimation in a negative-binomial model [Versión electrónica]. *Annals of the Institute of Statistical Mathematics*, 49(1), 101-115.
- Navarro, A., Utzet, F., Puig, P., Caminal, J. y Martín, M. (2001). La distribución binomial negativa frente a la de Poisson en el análisis de fenómenos recurrentes [Versión electrónica]. *Gaceta Sanitaria*, 15(5), 447-452.
- Nelder, J. A. y Wedderburn, W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society-Series A*, 135, 370-384.
- Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates [Versión electrónica]. *Journal of Quantitative Criminology*, 16(1), 21-43.
- Poortema, K. (1999). On modelling overdispersion of counts. *Statistica Neerlandica*, 53(1), 5-20.
- Reese, R. A. (1986). Data analysis: The need for models? *The Statistician*, 35, 199-206.
- Ribeiro, F. F. (1999). Underreporting of purchases of port wine [Versión electrónica]. *Journal of Applied Statistics*, 26(4), 185-494.
- Rodrigo, M. F. (2000). *Proyecto Docente de la Asignatura Análisis de Datos*. (Manuscrito no publicado). Valencia: Universitat de València. Departament de Metodologia de les Ciències del Comportament.
- Rodríguez, G. (2002). *Lecture Notes on Generalized Linear Models*. Recuperado el 22/2/2002 de <http://data.princeton.edu/wws509/notes>.
- Rothenberg, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica*, 52, 827-842.
- Saha, A. y Dong, D. (1997). Estimating nested count data models [Versión electrónica]. *Oxford Bulletin of Economics and Statistics*, 59(3), 423-430.
- Scollnik, D. P. M. (1995). Bayesian analysis of two overdispersed Poisson models [Versión electrónica]. *Biometrics*, 51, 1117-1126.
- Sejnowski, T. J., Koch, C. y Churchland, P. S. (1990). Computational neuroscience. En S.J.Hanson y C. R. Olson (Eds.), *Connectionist Modelling*

- and Brain Function: The Developing Interface* (pp. 5-35). Cambridge, MA: The MIT Press.
- Shankar, V., Milton, J. y Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry [Versión electrónica]. *Accident Analysis and Prevention*, 29(6), 829-837.
- Sturman, M. C. (1999). Multiple approaches to analyzing count data in studies of individual differences: The propensity for Type I errors, illustrated with the case of absenteeism prediction. *Educational and Psychological Measurement*, 59(3), 414-430.
- Suárez, C. P. (1996). Neurociencia y computación neuronal: una perspectiva. En C.P.Suárez y J. Regidor (Eds.), *Neurociencia y computación neuronal* (pp. 251-266). Las Palmas de Gran Canaria: Universidad de las Palmas de Gran Canaria. Sociedad Canaria de Neurociencia.
- Trivedi, P. K. (1997). Econometric models of event counts [Versión electrónica]. *Journal of Applied Econometrics*, 12, 199-201.
- Tu, W. y Zhou, X. (1999). A wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in Medicine*, 18, 2749-2761.
- van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, 51, 738-743.
- Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-334.
- Wang, W. y Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression [Versión electrónica]. *Journal of Population Economics*, 10, 273-283.
- Winkelmann, R. (1995). Duration dependence and dispersion in count data models. *Journal of Business and Economic Statistics*, 13(4), 467-474.
- Winkelmann, R. (2000). *Econometric Analysis of Count Data*. (3ª ed.). Berlin: Springer-Verlag.
- Winkelmann, R. y Zimmermann, K. F. (1991). A new approach for modeling economic count data. *Economic Letters*, 37, 139-143.
- Winkelmann, R. y Zimmermann, K. F. (1995). Recent developments in count data modelling: theory and application. *Journal of Economic Surveys*, 9(1), 1-24.

- Xekalaki, E. (1983). The univariate generalised Warning distribution in relation to accident theory: proneness, spells or contagion. *Biometrics*, 39, 887-895.
- Yau, K. K. W. y Lee, A. H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury programme [Versión electrónica]. *Statistics in Medicine*, 20, 2907-2920.
- Yen, S. T. (1999). Gaussian versus count-data hurdle models: cigarette consumption by women in the US [Versión electrónica]. *Applied Economic Letters*, 6, 73-76.