ESCUELA DE ADMINISTRACIÓN
FACULTAD DE ECONOMÍA Y ADMINISTRACIÓN

Tarea 1 Teoría Econométrica

Nombre del alumno:

Francisco Ignacio Fuentes Toro

Pregunta 1.2

Table 1:

	Dependent variable:
	C
Y	0.900***
	(0.000)
Constant	0.000
	(0.000)
Observations	1,000
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.000 (df = 998)
F Statistic	54,654,854,968,526,405,144,480,864,428.000*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Está estimación se basa en componentes completamente determinísticos, es por esto que la incertidumbre relacionada a los resultados es nula. El estimador de β es exactamente 0.9. Su intervalo de confianza no es un intervalo, en realidad es 0.9. El R^2 de la regresión es 1, por lo mismo es que el estadístico F tiende a infinito.

Pregunta 1.3

Table 2:

	Dependent variable:
	C
Y1	0.890***
	(0.885, 0.896)
Constant	95.651***
	(40.176, 151.126)
Observations	1,000
R^2	0.990
Adjusted R^2	0.990
Residual Std. Error	4.656 (df = 998)
F Statistic	98,961.380*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Ahora los pares de puntos que usamos para hacer esta regresión no siguen una proporción fija entre ellos, es más bien aproximada. En otras palabras, simulamos incertidumbre en los datos. Esto nos conduce a un estimador de β distinto de 0.9, con intervalo de confianza que ni siquiera contiene dicho valor (Par de valores encerrados entre paréntesis justo debajo del estimador). El R^2 también se ve afectado, definitivamente tendría que ser distinto de 1 en este ejercicio. Aunque el modelo sigue siendo significativo de acuerdo al test F.

Pregunta 1.4

Table 3:

	Dependent variable:
	C
Y2	0.890***
	(0.885, 0.896)
Constant	6.607
	(−49.423, 62.637)
Observations	1,000
R^2	0.990
Adjusted R^2	0.990
Residual Std. Error	4.656 (df = 998)
F Statistic	98,961.380*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notar que no cambia absolutamente nada más que el estimador del coeficiente de posición, lo que tiene mucho sentido puesto que literalmente añadimos un componente determinístico a los regresores sumándoles a todos 100. El nivel de incertidumbre se mantiene constante si a todas las variables les sumamos o le restamos el mismo valor, así que los resultados son intuitivos. Evidentemente el estimador de la constante que acompaña la recta disminuye considerando que corrimos todos los regresores a la derecha.

Pregunta 1.5

Table 4:

	Dependent variable:
	C1
Y	0.949***
	(0.887, 1.011)
Constant	−491.111
	(−1,108.461, 126.240)
Observations	1,000
R^2	0.476
Adjusted R^2	0.476
Residual Std. Error	51.518 (df = 998)
F Statistic	908.029*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

En esta iteración regresamos al modelo planteado originalmente, pero añadimos la incertidumbre en las variables del lado izquierdo. Sin embargo, esta vez el nivel de la varianza es mucho mayor, pasamos de 5 a 50, es 10 veces más grande. Esto se ve reflejado inmediatamente en intervalos de confianza considerablemente más anchos y es claro que los estimadores no son los valores que prestablecimos. El R^2 es el estadístico más afectado, pasando de ser cercano a 1 a prácticamente la mitad. A pesar de todo lo mencionado, el modelo sigue siendo estadísticamente significativo.

Pregunta 1.6

Table 5:

	Dependent variable:
	C2
Y	0.900***
	(0.900, 0.900)
Z	−0.250***
	(−0.250, −0.250)
Constant	0.000
	(−0.000, 0.000)
Observations	1,000
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.000 (df = 997)
F Statistic	6,602,726,207,762,575,482,606,260,846.000*** (df = 2; 997)
Note:	*p<0.1; **p<0.05; ***p<0.01

En esta iteración pasamos a un modelo multivariado, como veremos más adelante es para estudiar interacciones entre regresores. En este primer escenario nuevamente estamos frente a la ausencia de incertidumbre lo que se ve reflejado en los intervalos de confianza, el estadístico F y el R^2. Los estimadores son perfectos e idénticos a los valores prestablecidos.

Pregunta 1.7

Table 6:

	Dependent variable:
	C3
Y	0.949***
	(0.887, 1.011)
Z	−0.231***
	(−0.354, −0.108)
Constant	−491.728
	(−1,109.373, 125.918)
Observations	1,000
R^2	0.479
Adjusted R^2	0.478
Residual Std. Error	51.542 (df = 997)
F Statistic	457.558*** (df = 2; 997)
Note:	*p<0.1; **p<0.05; ***p<0.01

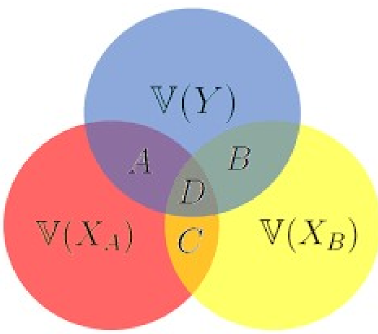
Ahora agregamos incertidumbre al modelo anterior, eso lo hace más divertido. Lo primero en lo que tenemos que percatarnos es que los estimadores son distintos de los parámetros prestablecidos, aunque sus verdaderos valores están dentro de los intervalos de confianza. El R^2 es propio del nivel de incertidumbre que hemos agregado. Sin embargo, el modelo sigue siendo significativo bajo el test F.

Pregunta 1.8

Table 7:

	Dependent variable:	
	C3	
	(1)	(2)
Z	−0.187**	
	(−0.357, −0.017)	
Y		0.946***
		(0.884, 1.008)
Constant	8,990.721***	−499.271
	(8,964.775, 9,016.667)	(−1,120.770, 122.227)
Observations	1,000	1,000
R^2	0.005	0.472
Adjusted R^2	0.004	0.471
Residual Std. Error (df = 998)	71.178	51.864
F Statistic (df = 1; 998)	4.635**	890.400***
Note:		*p<0.1; **p<0.05; ***p<0.01

Ahora tenemos dos regresiones separadas cuyos estimadores no cambian en demasía de la pregunta anterior estadísticamente hablando. Esto se debe a que la correlación entre las variables Z e Y no es mucha, es más, para ser exactos es "0.02347154". Esto es consecuencia directa del teorema de Frish Waugh.



Notar que, como se muestra en este diagrama explicativo, la incidencia que tienen las variables X_A y X_B sobre Y depende tanto de sus efectos directos, A y B respectivamente, como de su efecto conjunto D. En este caso especifico Z e Y no tienen un efecto conjunto muy considerable puesto que no guardan una relación significativa. En otras palabras "tienen un D pequeño".

Pregunta 1.9

Table 8:

	Dependent variable:
	C4
Y	1.234***
	(0.680, 1.787)
W	−0.642**
	(−1.256, −0.028)
Constant	195.448
	(−391.808, 782.703)
Observations	1,000
R^2	0.329
Adjusted R^2	0.327
Residual Std. Error	48.949 (df = 997)
F Statistic	243.886*** (df = 2; 997)
Note:	*p<0.1; **p<0.05; ***p<0.01

Ahora tenemos un modelo similar al de la pregunta 7, pero con una gran diferencia en cuanto a los regresores. Los coeficientes parecen alejarse aún más de sus valores prestablecidos y sus intervalos de confianza presentan una varianza más grande puesto que son notablemente más anchos que antes. El R^2 es incluso menor que en el modelo establecido en la pregunta 7, sin embargo, el test F denota que todavía existe una significancia estadística considerable.

Pregunta 1.10

Table 9:

	Dependent variable:	
	C4	
	(1)	(2)
W	0.719***	
	(0.653, 0.784)	
Y		0.658***
		(0.599, 0.717)
Constant	270.682	165.582
	(−320.889, 862.253)	(−421.918, 753.081)
Observations	1,000	1,000
R^2	0.316	0.326
Adjusted R^2	0.315	0.325
Residual Std. Error (df = 998)	49.390	49.027
F Statistic (df = 1; 998)	460.351***	482.027***
Note:		*p<0.1; **p<0.05; ***p<0.01

La situación ahora es completamente opuesta a la presenciada en la pregunta 8. Puesto que la correlación entre W y Z es muy cercana a 1, para ser exactos es de "0.9943686". La lógica aplica de la misma forma que antes, solo que ahora el impacto de "D" es considerablemente mayor, lo que hace que hacer ambas estimaciones por separado de resultados radicalmente distintos de la regresión conjunta. Se subestima el valor del coeficiente de Y y se sobrestima el valor del coeficiente de W.

Pregunta 1.11

Table 10:

	Dependent variable:
	C3
Y	0.900***
	(0.898, 0.902)
Z	−0.232***
	(−0.355, −0.109)
Observations	1,000
R^2	0.4778372
Residual Std. Error	51.579 (df = 998)
F Statistic	456.641*** (df = 2; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notar que los estimadores cambiaron levemente, sin embargo, el de Y fue el más afectado acercándose a su valor prestablecido y reduciendo el ancho de su intervalo de confianza significativamente. A pesar de que los resultados parecen ser convenientes, el R^2 disminuyó, lo que es lógico pues le arrebatamos una variable de decisión al modelo. El test F no cambia no cambia en gran medida y el modelo sigue siendo estadísticamente significativo en su totalidad.

Pregunta 1.12

Table 11:

	Dependent variable:		
	Cm		C1m
	(1)	(2)	(3)
Ym	0.900***		0.816***
	(0.900, 0.900)		(0.503, 1.128)
Y1m		0.893***	
		(0.867, 0.919)	
Constant	−0.000	68.529	839.333
	(−0.000, 0.000)	(−188.613, 325.671)	(−2,283.776, 3,962.441)
Observations	50	50	50
R^2	1.000	0.990	0.353
Adjusted R^2	1.000	0.990	0.339
Residual Std. Error (df = 48)	0.000	4.923	59.638
F Statistic (df = 1; 48)	3,436...***	4,628.817***	26.181***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 12:

	Dependent variable:	
	C3m	C4m
	(1)	(2)
Ym	0.835***	0.091
	(0.511, 1.159)	(−2.788, 2.970)
Zm	−0.055	
	(−0.831, 0.721)	
Wm		0.595
		(−2.569, 3.758)
Constant	617.184	484.276
	(−2,652.462, 3,886.830)	(−2,103.912, 3,072.464)
Observations	50	50
R^2	0.368	0.328
Adjusted R^2	0.341	0.300
Residual Std. Error (df = 47)	60.114	49.313
F Statistic (df = 2; 47)	13.688***	11.484***
Note:		*p<0.1; **p<0.05; ***p<0.01

Notar que el cambio más significativo reside en un aumento de la varianza estimada de todos los coeficientes estimados exceptuando aquel bajo completa certidumbre claramente. Esto se observa por medio de verificar que todos los intervalos de confianza son considerablemente más anchos que en las preguntas anteriores. Sin embargo, todos los modelos siguen siendo estadísticamente significativos bajo el test F.

Pregunta 2.1 Cambiando el tamaño de la muestra

Notar que al cambiar el tamaño de la muestra, la forma de la distribución no parece variar demasiado. Pero la varianza de cada estimador OLS disminuye considerablemente cada vez que aumentamos el tamaño. Esto en consistente con lo estudiado en clases puesto que la varianza de los estimadores esta multiplicada por la matriz inversa del producto de la matriz de regresores multiplicada con si misma, termino que crece a medida que aumentamos más valores en la muestra. Ocurre lo mismo para el estadístico F, los casos extremos son cada vez más extraños hasta el punto de que prácticamente ya no se encuentra ninguno luego del valor 3 con m=500.

Pregunta 2.2 Cambiando el número de repeticiones

Aquí apreciamos claramente en acción el teorema del límite centra de Lindberg Levy. Cada vez que las repeticiones del experimento aumentan los coeficientes estimados van tomando la forma de una distribución normal. De manera similar ocurre con la distribución del estadístico F, cada iteración nos muestra como toma la forma característica de una distribución Fisher.

Pregunta 2.3 Cambiando el origen de los errores

La distribución de los errores parece no cambiar en gran medida el escenario. Sin embargo, el principal afectado es el estimador del coeficiente de posición, el cual parece ser mayor que el estimado con errores normales tanto con los errores poisson como los uniformes. Este cambio parece ser consecuencia directa de que los errores son todos no negativos en dichos escenarios, esto se debe a la naturaleza de las distribuciones. Es decir, una distribución uniforme entre 0 y 1, o una poisson jamás tomarán valores negativos, dicha característica parece incidir directamente en el estimador del coeficiente de posición.

Figura 1 Cambiando el tamaño de muestra

1000 repeticiones, errores normales

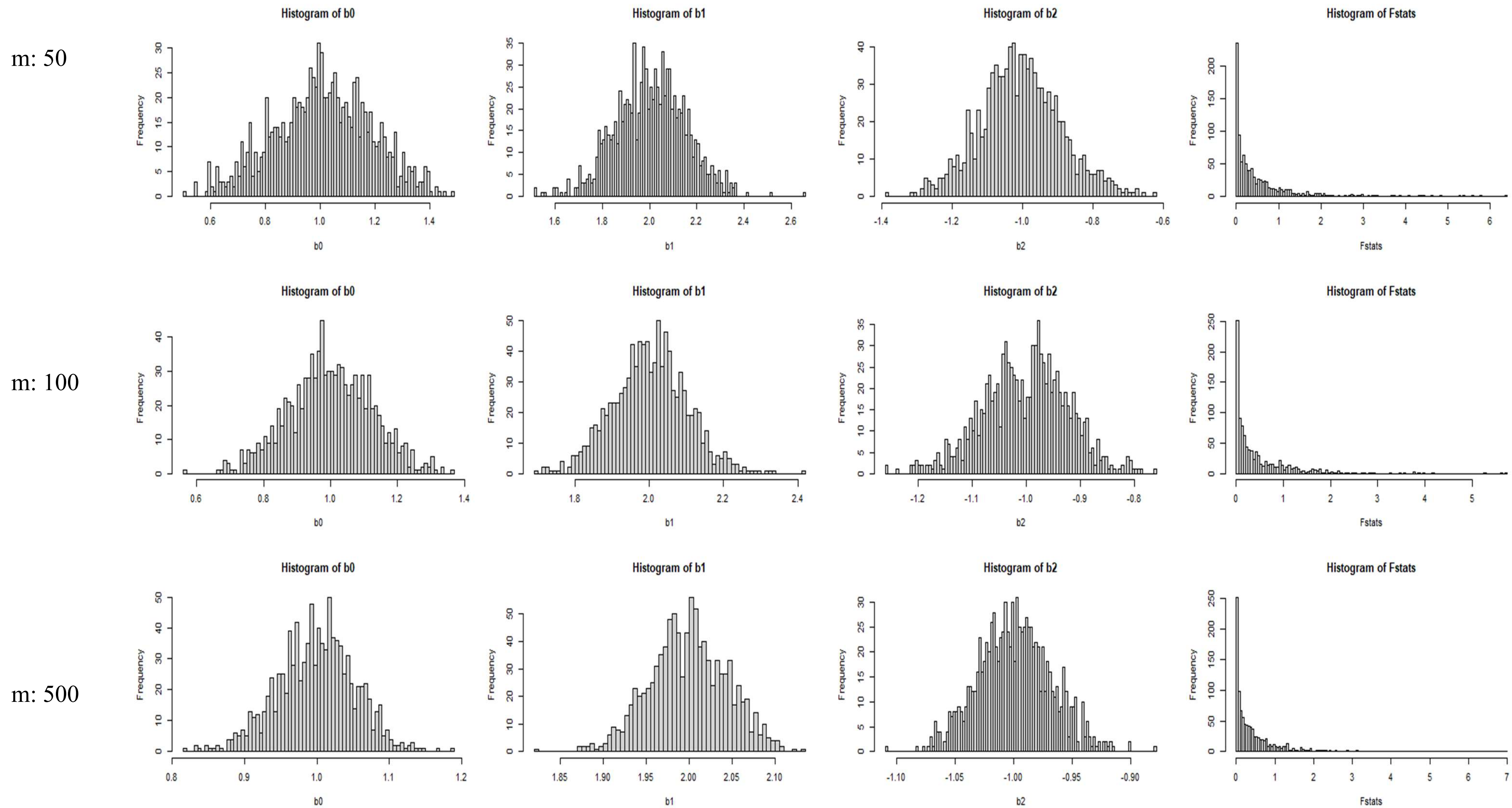


Figura 2 Cambiando el número de repeticiones

1000 elementos en la muestra, errores normales

Rep: 50

Rep: 100

Rep: 500

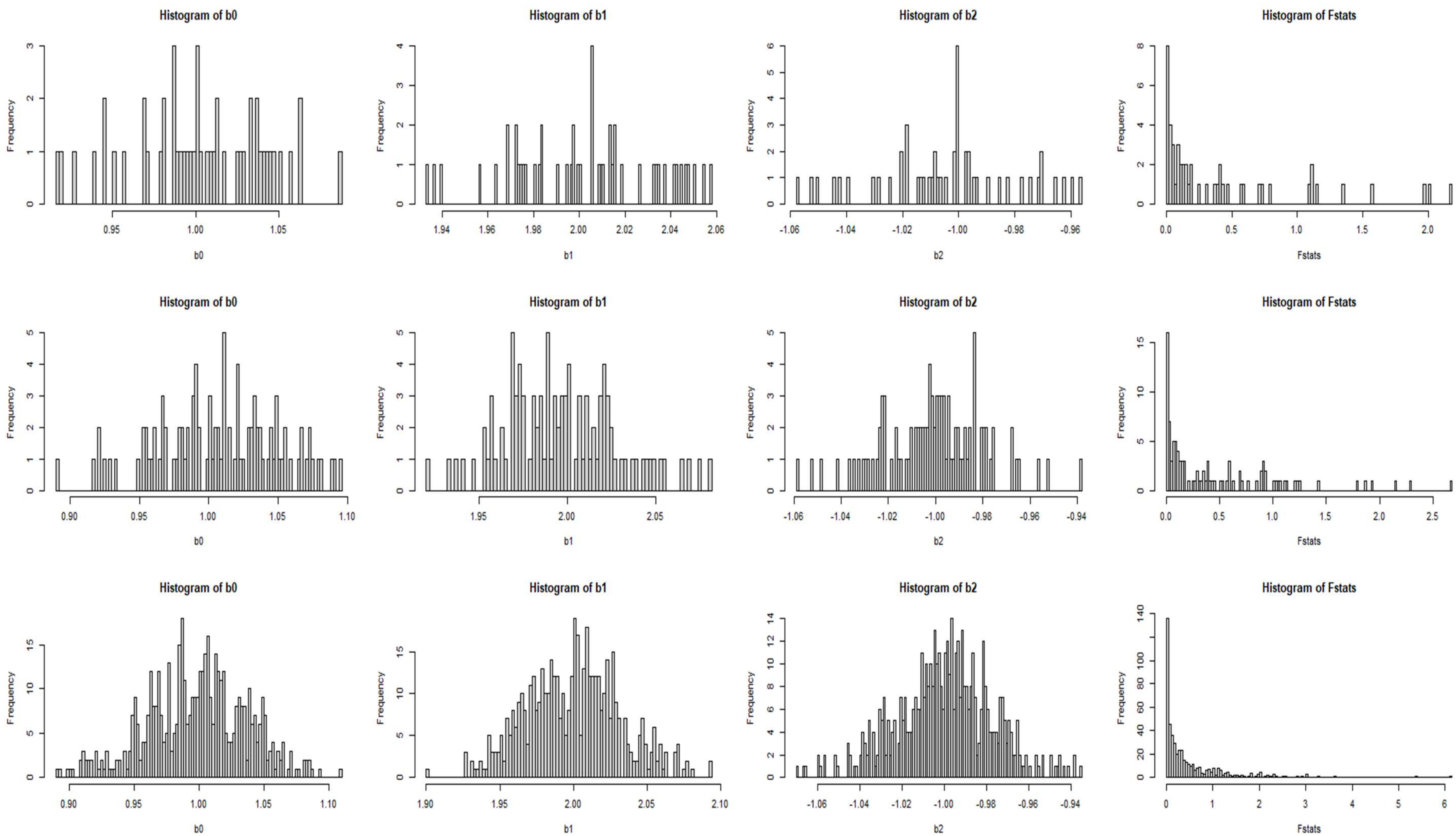


Figura 3 Cambiando la distribución del error

1000 elementos en la muestra y 100 elementos en cada muestra

Errores normales

Errores Uniformes entre 0 y 1

Errores Poisson