

## TEST DE RAZÓN DE VEROSIMILITUD GENERALIZADA

JUAN C. MANRÍQUEZ<sup>1</sup>, LUIS CID<sup>2</sup> y MARCELA VALDEZ<sup>3</sup>

1 *Universidad del Bío – Bío, Facultad de Ciencias, Departamento de Matemáticas*  
[jcmanriq@ubiobio.cl](mailto:jcmanriq@ubiobio.cl)

2 *Universidad del Bío – Bío, Facultad de Ciencias, Departamento de Estadística*  
[lcid@ubiobio.cl](mailto:lcid@ubiobio.cl)

3 *Universidad del Desarrollo, Facultad de Odontología.* [marcelavaldes@udd.cl](mailto:marcelavaldes@udd.cl)

### RESUMEN

Este trabajo tiene por objetivo estudiar las condiciones de convergencia del Test de Razón de Verosimilitud Generalizada (TRVG), para probar hipótesis del tipo  $H_0: \mu = \mu_0$  v/s  $H_a: \mu \neq \mu_0$ . En lo particular, estudiar el comportamiento de los errores tipo I y tipo II, bajo distintas parametrizaciones y tamaños muestrales.

Se analizan separadamente distintos tamaños muestrales y los resultados de usar el TRVG, se comparan con los obtenidos al utilizar aproximaciones obtenidas, por ejemplo asumiendo las condiciones para aplicar el Teorema Central del Límite.

**PALABRAS CLAVE:** Error tipo I y II, test de razón de verosimilitud, teorema central del límite.

## 1. INTRODUCCIÓN

La calidad y tamaño de la muestra juegan un papel de primera importancia, al momento de medir la cantidad de evidencia que existe en contra o a favor de alguna hipótesis. Sin lugar a dudas, otro elemento que influye es el tipo de test que se utiliza. Si consideramos por ejemplo, el problema de probar hipótesis del tipo

$$H_0: \mu = \mu_0 \quad \text{v/s} \quad H_a: \mu \neq \mu_0 \quad (1)$$

En que  $\mu$  corresponde a la media de una población con distribución normal con parámetros  $\mu$  y  $\sigma^2$ , ambos no necesariamente conocidos; la práctica estadística corriente recomienda, bajo ciertas condiciones de regularidad, el uso del test T, basado en la media muestral. Este test es fácilmente derivable como un test de razón de verosimilitud generalizado, bajo condiciones de normalidad. Sin embargo, en casos en que la distribución del estadístico resultante de la razón de verosimilitud, no tiene distribución conocida, ésta tiene distribución asintótica chi-cuadrado.

El objetivo de este trabajo es analizar la convergencia del test de razón de verosimilitud generalizado, bajo condiciones de normalidad, comparándolo con el test T común, que utiliza resultados de convergencia basados en el Teorema Central del Límite, llamaremos a estos tests  $T_2(x)$  y  $T_1(x)$  respectivamente. En particular estamos interesados en analizar el comportamiento de los errores Tipo I y Tipo II, bajo diferentes condiciones de tamaño muestral y varianza de la distribución.

## 2. METODOLOGÍA

### *Estadístico $T_1(X)$ .*

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con función de densidad de probabilidad conocida, entonces el test estadístico, se puede verificar que el TRVG, tiene la forma

$$T_1(\mathbf{x}) = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim t_{(1-\frac{\alpha}{2}, n-1)} \quad (1)$$

Este test tiene región crítica de tamaño  $\alpha$ , definida por

$$W_1 = \left\{ x: T_1(\mathbf{x}) > t_{\left(1-\frac{\alpha}{2}, n-1\right)} \right\} \quad (2)$$

En consecuencia, el Error tipo I  $= \alpha = \Pr (T_1(\mathbf{x}) \in W_1 | H_0)$ , y a su vez el Error tipo II queda dado por  $\beta = \Pr (T_1(\mathbf{x}) \notin W_1 | H_a)$

***Estadístico  $T_2(X)$ .***

El TRVG es un caso más general del Lema de Neyman – Pearson, definido para el caso de hipótesis simples (Bartoszyński y Niewiadomska 2008) y que para el caso de hipótesis compuestas deriva en pruebas uniformemente más potentes similares. En efecto, de acuerdo con este Lema, para una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una población con densidad de probabilidades  $f(x, \theta)$ , si deseamos probar la hipótesis

$$H_0: \theta = \theta_0 \quad \text{v/s} \quad H_a: \theta = \theta_a$$

El Test de Razón de Verosimilitud, define una región crítica (RC)  $W_L$  tal que

$W_L = \{ \underline{x}: \lambda = \frac{f_a(\theta|x)}{f_0(\theta|x)} > k, \quad k > 0 \}$  es de tamaño  $\alpha$  (fijo). Lo cual genera una región crítica para  $\bar{x}$ , dada por  $W = \{ \underline{x}: \bar{x} \geq k \}$ .

Una extensión del método de razón de verosimilitud que a menudo conduce a un buen desempeño de las pruebas es el Test de Razón de Verosimilitud Generalizada (TRVG) y se usa cuando estamos interesados en probar las hipótesis acerca de un parámetro específico, pero la distribución de la población depende también de algún otro parámetro.

**Test de Razón de Verosimilitud Generalizada:**

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con función de densidad de probabilidad multiparamétrica dada por  $f(x, \theta)$ , con  $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ . Si  $\theta \in \Theta_0$ , en que  $\Theta_0$  es el correspondiente espacio paramétrico bajo  $H_0$  (Bartoszyński y Niewiadomska 2008).

Entonces, el estadístico  $T_2(\mathbf{X}) = -2 \log v(\mathbf{X})$  tiene distribución asintótica  $\chi_{m-r}^2$ , donde  $r$  es el número de parámetros de  $\theta$  que ha sido completamente especificado en la hipótesis nula y  $v(x)$  tiene la siguiente expresión:

$$v(\mathbf{X}) = \frac{\sup_{\theta \in \Theta_0} f_n(x, \theta)}{\sup_{\theta \in \Theta} f_n(x, \theta)} = \frac{\sup L_0}{\sup L} = \frac{f_n(x, \hat{\theta}_0)}{f_n(x, \hat{\theta})} \quad (3)$$

Dado que este trabajo consideró una población cuya distribución fue  $N(\mu, \sigma^2)$ , la ecuación (3) se resume a la siguiente:

$$v(\mathbf{X}) = \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^{-n/2} \quad (4)$$

De esta manera, el estadístico de prueba  $T_2(\mathbf{X})$  queda la forma

$$T_2(\mathbf{x}) = -2 \log[v(\mathbf{X})] = -2 \log \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^{-n/2} \sim \chi_{m-r}^2 \quad (5)$$

Esta ecuación (5) nos indica que a medida que  $n$  aumenta, la distribución  $-2 \log v(\mathbf{X})$ , bajo hipótesis nula, tiende a una distribución chi – cuadrado con  $m - r$  grados de libertad (Kendall y Stuart 1960)

Luego el test que define la Región Crítica (RC):

$$W_2 = \{x: T_2(\mathbf{x}) > \chi_{m-r}^2\} \quad (6)$$

En consecuencia, el Error tipo I =  $\Pr (T_2(\mathbf{x}) \in W_2 | H_0)$  y a su vez el Error tipo II queda dado por  $\beta = \Pr (T_2(\mathbf{x}) \notin W_2 | H_a)$ .

***Estudio de simulación.***

Este estudio considera la comparación de los errores tipo I y tipo II, asociados a los estadísticos de prueba indicados en (1) y (5), cuando varían los tamaños muestrales y las varianzas de la población. Para ello, se realizó un estudio de simulación, mediante el cual se generaron, utilizando rutinas programadas en R, muestras de números aleatorios provenientes de una distribución normal, para distintos tamaños de muestras y considerando diferentes varianzas.

Los tamaños muestrales empleados fueron  $n=10, 20, 30, 60, 100$  y  $200$ . Se trabajó con una única hipótesis nula. Esto es,  $H_0: \mu = \mu_0 = 350$  vs.  $H_a: \mu \neq 350$ .

En que el valor 350 fue elegido arbitrariamente.

Las varianzas fueron seleccionadas de modo que las desviaciones estándar tomaron valores de 50, 100, 200, 300. Se generaron 200 muestras para cada combinación de tamaño muestral y desviación estándar (24 combinaciones en total).

***Error tipo I***

El error tipo I se define como probabilidad (denotado por  $\alpha$ ) de rechazar una hipótesis ( $H_0$ ), cuando ésta no debiera ser rechazada; esto es,

$$\alpha = Prob(\bar{x} \in RC / \mu = 350)$$

A continuación, se muestra la metodología a seguir para evaluar cada tipo de error. Como ya hemos señalado, los estadísticos de prueba a usar en cada caso son los siguientes:

i) Para  $T_1(\mathbf{x}) = \frac{\bar{X}-350}{\sigma/\sqrt{n}}$  (12)

Donde  $\bar{X}$  es la media de cada muestra generada,  $\sigma$  es la desviación estándar fijada (en nuestra simulación, 50, 100, 200 y 300) y  $n$  es el tamaño muestral (en esta simulación, 10, 20, 30, 60, 100 y 200). Se trabaja con un nivel de significación de 0.05. El valor de contraste luego será:  $t_{(.975, n-1)}$ .

ii) Para  $T_2(\mathbf{x})$ , se tiene:

$$T_2(\mathbf{x}) = -2 \cdot \log \left[ \frac{\sum_{i=1}^n (x_i - 350)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^{-\frac{n}{2}} \sim \chi_1^2 \quad (13)$$

Los grados de libertad de Chi-cuadrado resultan de la diferencia  $m - r = 2 - 1 = 1$ , con  $m$  el número de parámetros ( $\mu$  y  $\sigma^2$ ) y  $r$  el número de parámetros especificados en la prueba ( $\mu$ ). Para ambos casos (*i* y *ii*). Las muestras aleatorias fueron generadas mediante rutinas de R, para una población normal con media 350.

Para una de las 200 muestras de tamaño 10, se obtuvieron las respectivas medias muestrales, las que fueron comparadas con el valor propuesto en  $H_0$  obteniéndose los valores observados para el test correspondiente los que fueron comparados con los respectivos valores críticos. El procedimiento se realizó para cada una de las 200 muestras, obteniéndose así el porcentaje de rechazo de la hipótesis.

### ***Error tipo II***

El error tipo II se define como probabilidad (denotado por  $\beta$ ) de aceptar la hipótesis nula ( $H_0$ ), cuando ésta es falsa; esto es,

$$\beta = \text{Prob}(\bar{x} \notin RC / \mu \neq 350)$$

Para estimar el error tipo II, se generaron 200 muestras aleatorias de tamaños  $n=10$ ,  $n=20$ ;  $n=30$ ;  $n=60$ ;  $n=100$  y  $n=200$  provenientes de una población con distribución normal,  $\mu = 300$   $\mu = 350$ ;  $\mu = 360$ ;  $\mu = 370$ ;  $\mu = 380$ ;  $\mu = 390$  y  $\mu = 400$ , con  $\sigma = 50$ , con  $\sigma = 100$ ;  $\sigma = 200$  y  $\sigma = 300$ .

Se registraron los casos en que los valores observados de los estadísticos  $T_1$  y  $T_2$ , excedieron los valores críticos de tabla, para determinar el porcentaje de rechazo en cada una de las combinaciones. Los resultados fueron graficados.

### 3. RESULTADOS Y DISCUSIÓN

Se muestran a continuación los errores Tipo I y Tipo II obtenidos para ambos Test.

#### 3.1 Error tipo I

En base a las 200 muestras generadas para cada combinación de tamaño muestral y desviación estándar, se calculó el error tipo I que se muestra en la Tabla 1 siguiente.

Tabla 1. Comportamiento del error tipo I.

Desviación estándar	Estadístico	Tamaño muestral $n$					
		10	20	30	60	100	200
50	T	0,020	0,015	0,045	0,045	0,050	0,060
	TRVG	0,055	0,020	0,065	0,035	0,065	0,045
100	T	0,015	0,035	0,035	0,040	0,040	0,055
	TRVG	0,055	0,060	0,050	0,045	0,025	0,035
200	T	0,030	0,050	0,040	0,035	0,030	0,040
	TRVG	0,085	0,080	0,045	0,030	0,030	0,045
300	T	0,030	0,020	0,035	0,065	0,045	0,060
	TRVG	0,075	0,055	0,035	0,060	0,055	0,055

Como se observa, si mantenemos fija la desviación estándar, las veces que se rechaza aumentan para ambos estadísticos a medida que crece el valor de  $n$ . Sin embargo, para muestras inferiores a 60, el valor de  $\alpha$  es mayor o igual en la prueba TRVG que para T. Para muestras mayores o iguales a 60, tales magnitudes se invierten, es decir, el error tipo I se acrecienta bastante para el estadístico T con respecto al TRVG. Sin embargo, los test tienden a igualarse cuando la desviación estándar la aumentamos en los niveles que se indican. Para muestras de tamaño 200 el error tipo I que entrega el TRVG termina en torno a los valores 0,035 y 0,055; en tanto que el estadístico T oscila entre 0,040 y 0,060.

Las siguientes gráficas permiten tener una mejor apreciación del comportamiento de cada estadístico en torno al error tipo I.

Figura 1. Comportamiento del error tipo I, con  $\sigma = 50$

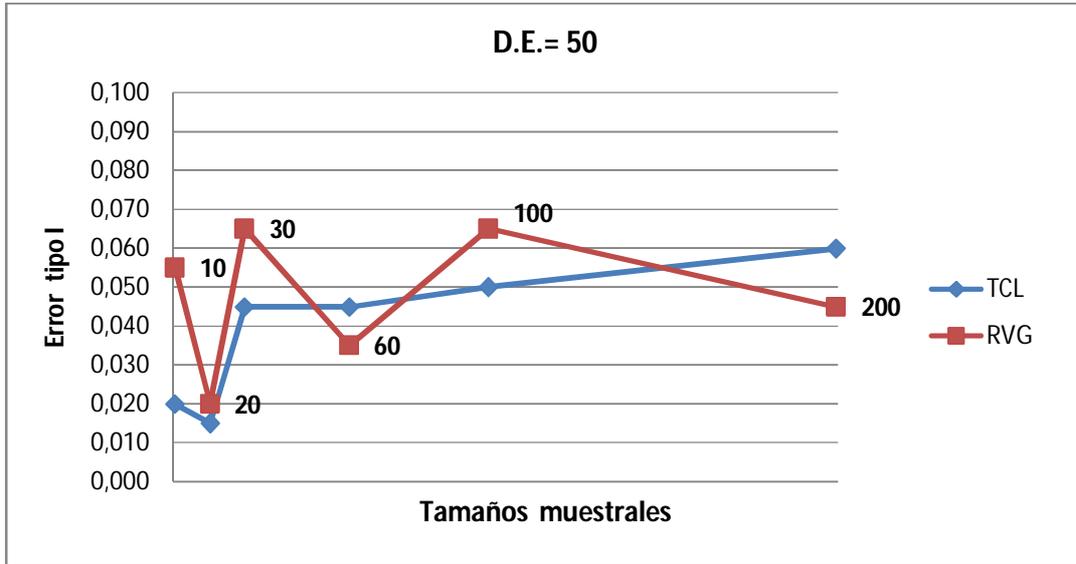


Figura 2. Comportamiento del error tipo I, con  $\sigma = 100$

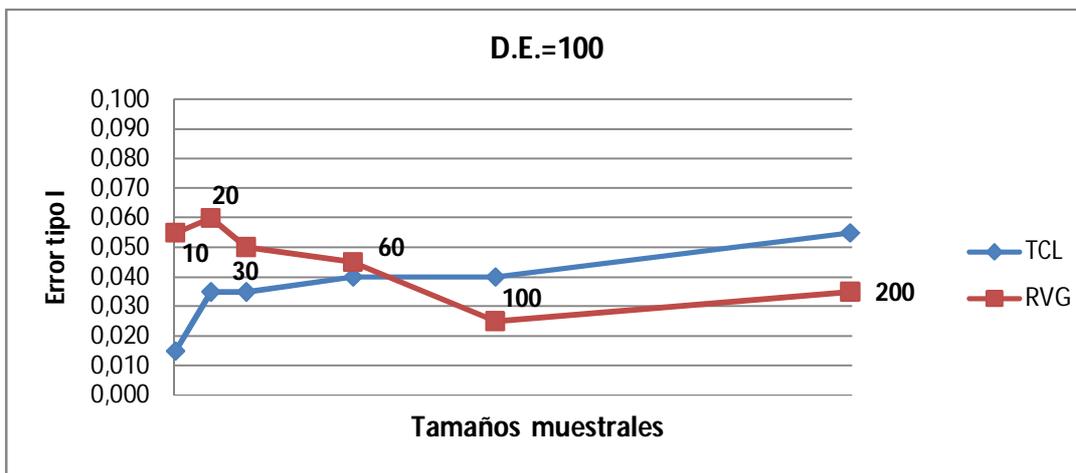


Figura 3. Comportamiento del error tipo I, con  $\sigma = 200$

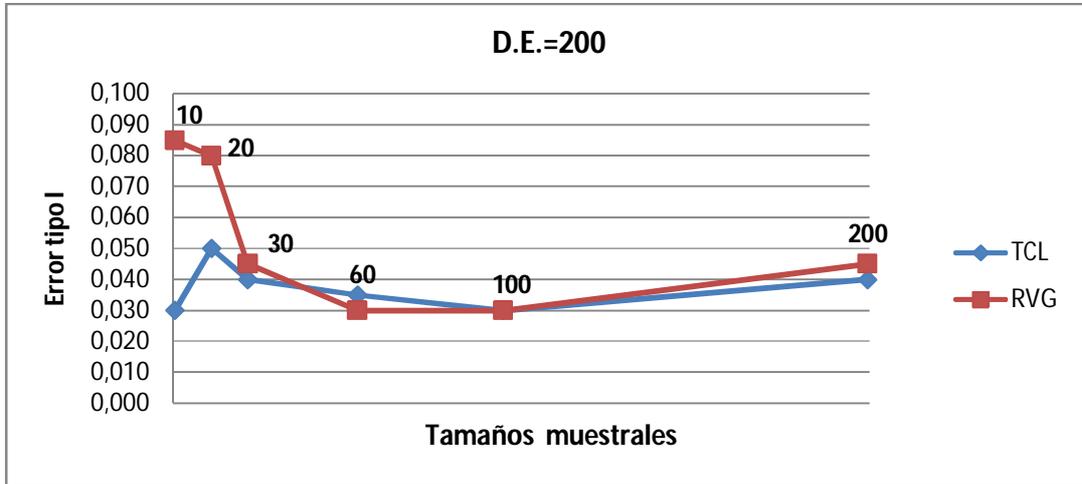
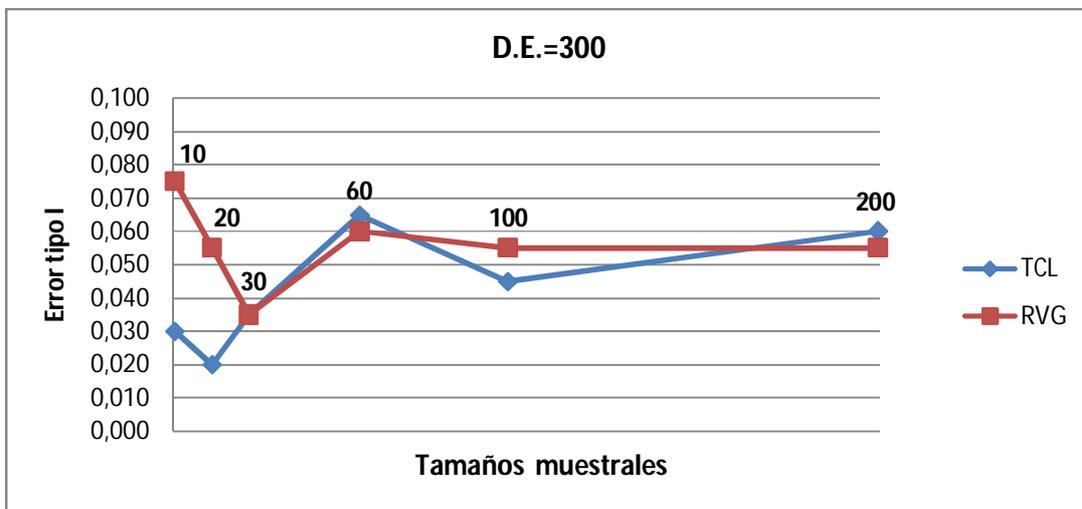


Figura 4. Comportamiento del error tipo I, con  $\sigma = 300$



Se debe hacer notar que para grandes desviaciones estándar y tamaños de muestras (100 y 200, en nuestras simulaciones) el estadístico del TRVG presenta menor error tipo I en la gran mayoría de los casos, que el estadístico  $T$ . En tamaños de muestra inferiores a 60 resulta lo contrario.

Otro aspecto importante a destacar, es que las variaciones que se observan en el error tipo uno, vienen dadas por el aumento en el tamaño de la muestra, más que la desviación estándar.

### 3.2 Error tipo II

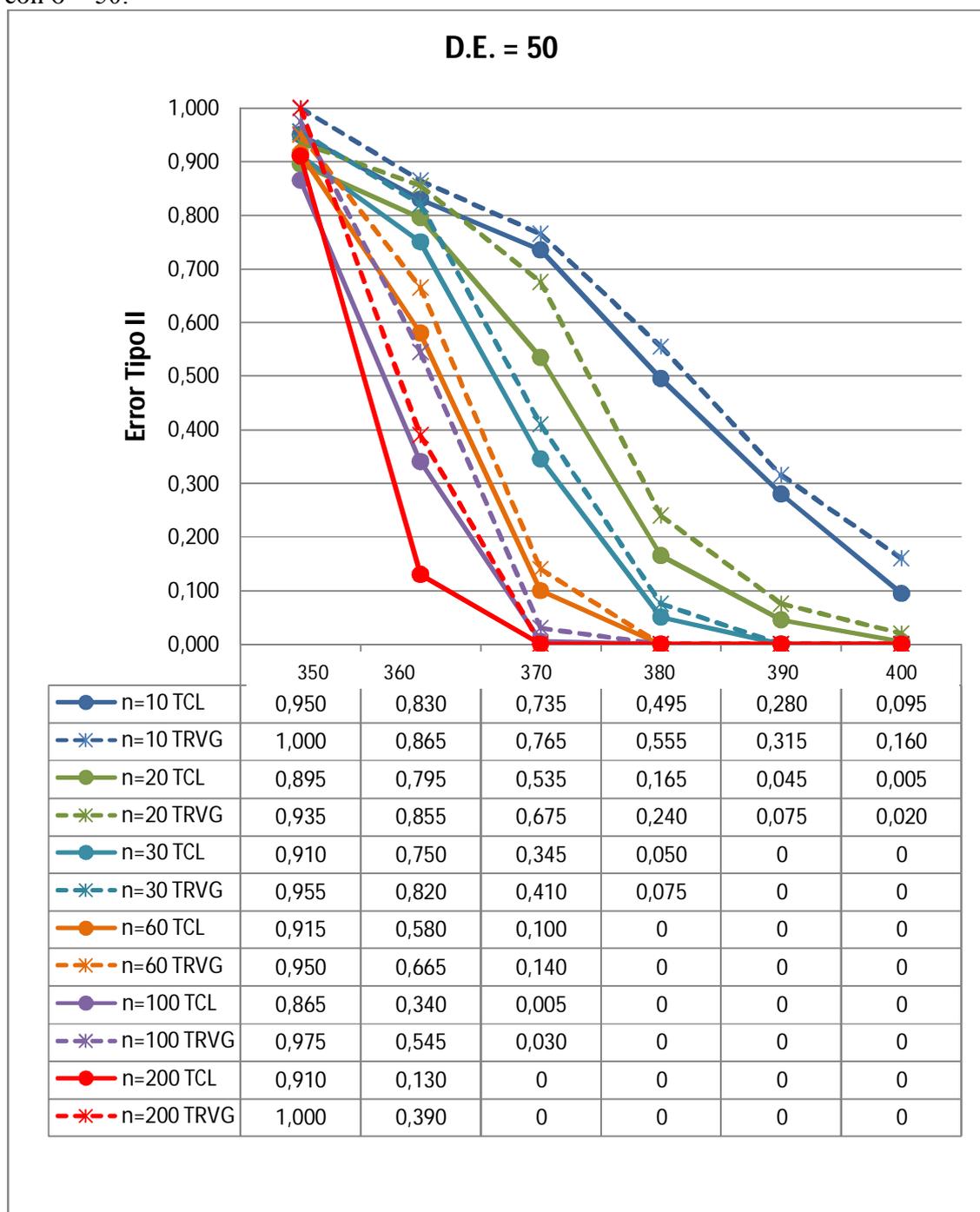
Para evaluar el error Tipo II de cada uno de los test en estudio, se procedió a graficar las curvas de este error en función de la media poblacional bajo  $H_1$ , de acuerdo al tamaño muestral y desviación estándar.

En la figura 5, se presenta la curva de error tipo II, en función de la media poblacional bajo la Hipótesis alternativa, para una desviación estándar de 50.

Se observa que a medida que la media bajo  $H_1$  está más cerca de la media bajo  $H_0$ , mayor es el error tipo II, independiente del tamaño de la muestra. Esto es lógico, ya que este error se refiere a la probabilidad de aceptar  $H_0$  dado que ésta es falsa, en otras palabras, este error se comete cuando el test no es capaz de detectar que la media poblacional es distinta a la propuesta; es de esperar que esto ocurra con mayor frecuencia cuando  $\mu_0$  y  $\mu_1$  son muy parecidas, y a medida que se van distanciando.

En cuanto al tamaño muestral, se observa que a medida que éste aumenta, el error tipo II disminuye drásticamente. Por ejemplo, para detectar una diferencia en la media bajo  $H_0 = 350$  y bajo  $H_1 = 370$ , con un tamaño muestral de 10 (y evaluando en un total de 200 muestras aleatorias) el error tipo II es de 0,75 aproximadamente con ambos Test, mientras que para muestras con la misma media, pero con un tamaño muestral de 200, este error disminuye a 0, lo que indica que en la totalidad de los casos, ambos Test fueron capaces de detectar la diferencia. Al comparar el error tipo II estimado con el estadístico  $T$  y con TRVG, se observa que para este último, es siempre mayor, para toda media poblacional y tamaño muestral.

Figura 5. Comportamiento del error tipo II, según media poblacional y tamaño muestral, con  $\sigma = 50$ .



En las figuras siguientes, se muestra el error tipo II, evaluado con desviaciones estándar mayores, en este caso, 100, 200 y 300. Se observa que los comentarios hechos para la figura anterior, también son válidos acá, puesto que se observa en todo punto que el estadístico T presenta menor nivel de error tipo II, y que éste disminuye también al momento de aumentar el tamaño muestral. Pero a medida que la varianza es mayor, se observa que es mayor la posibilidad de cometer error tipo II. Esto se explica, ya que el valor calculado para cada test, se ve afectado por la varianza de los datos, esto es, a mayor variabilidad, menor es el valor calculado, lo que produce que con mayor frecuencia se ubique erróneamente en la zona de aceptación de la hipótesis nula. Por ejemplo, en la figura 8, con una desviación estándar de 300 y con una media poblacional de 400, que está bastante alejada de la media bajo  $H_0$ , el error tipo II es muy alto, aún para un tamaño muestral considerado “grande” como lo es  $n=100$  o  $n=200$ .

Figura 6. Comportamiento del error tipo II, según media poblacional y tamaño muestral, con  $\sigma = 100$ .

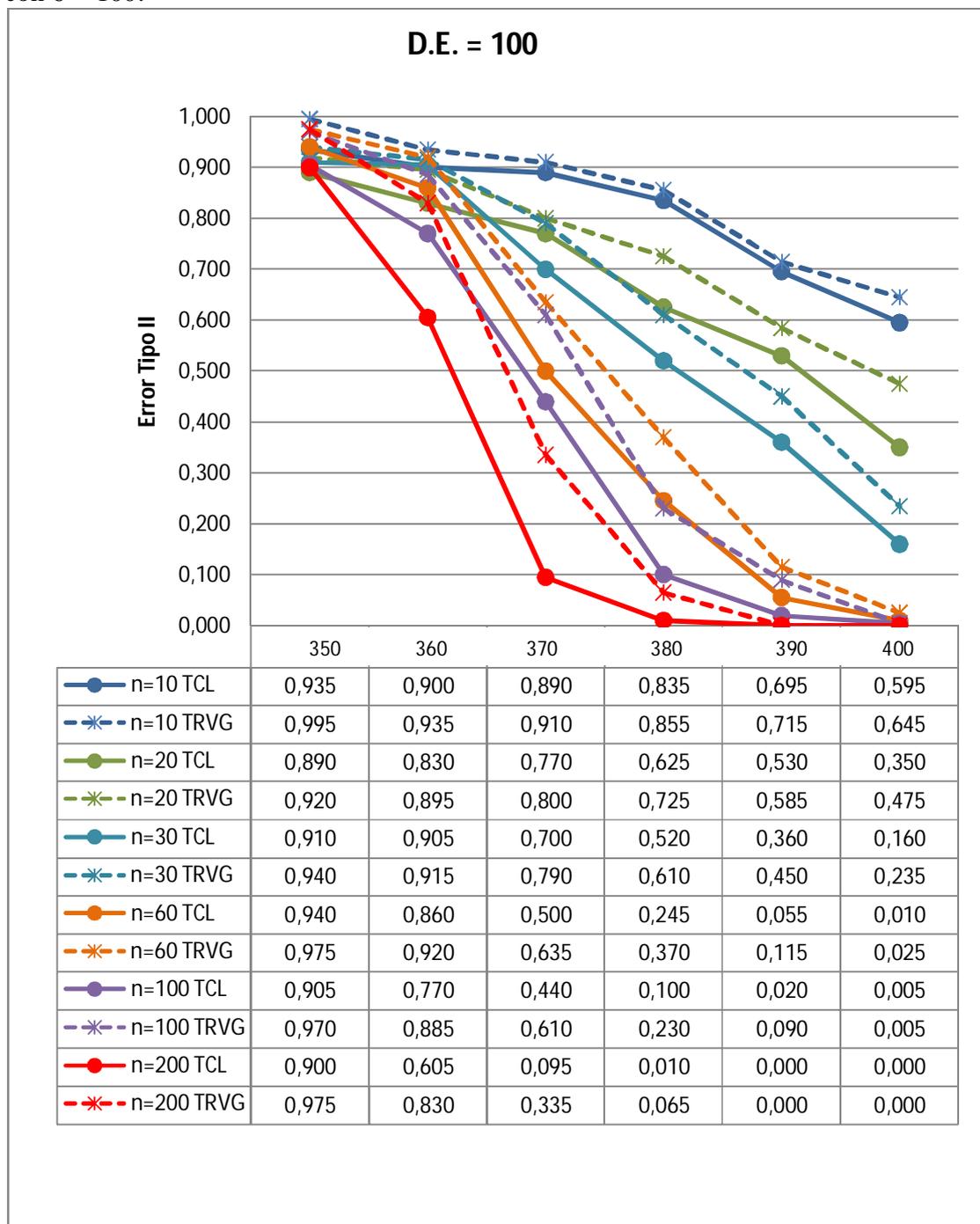


Figura 7. Comportamiento del error tipo II, según media poblacional y tamaño muestral, con  $\sigma = 200$ .

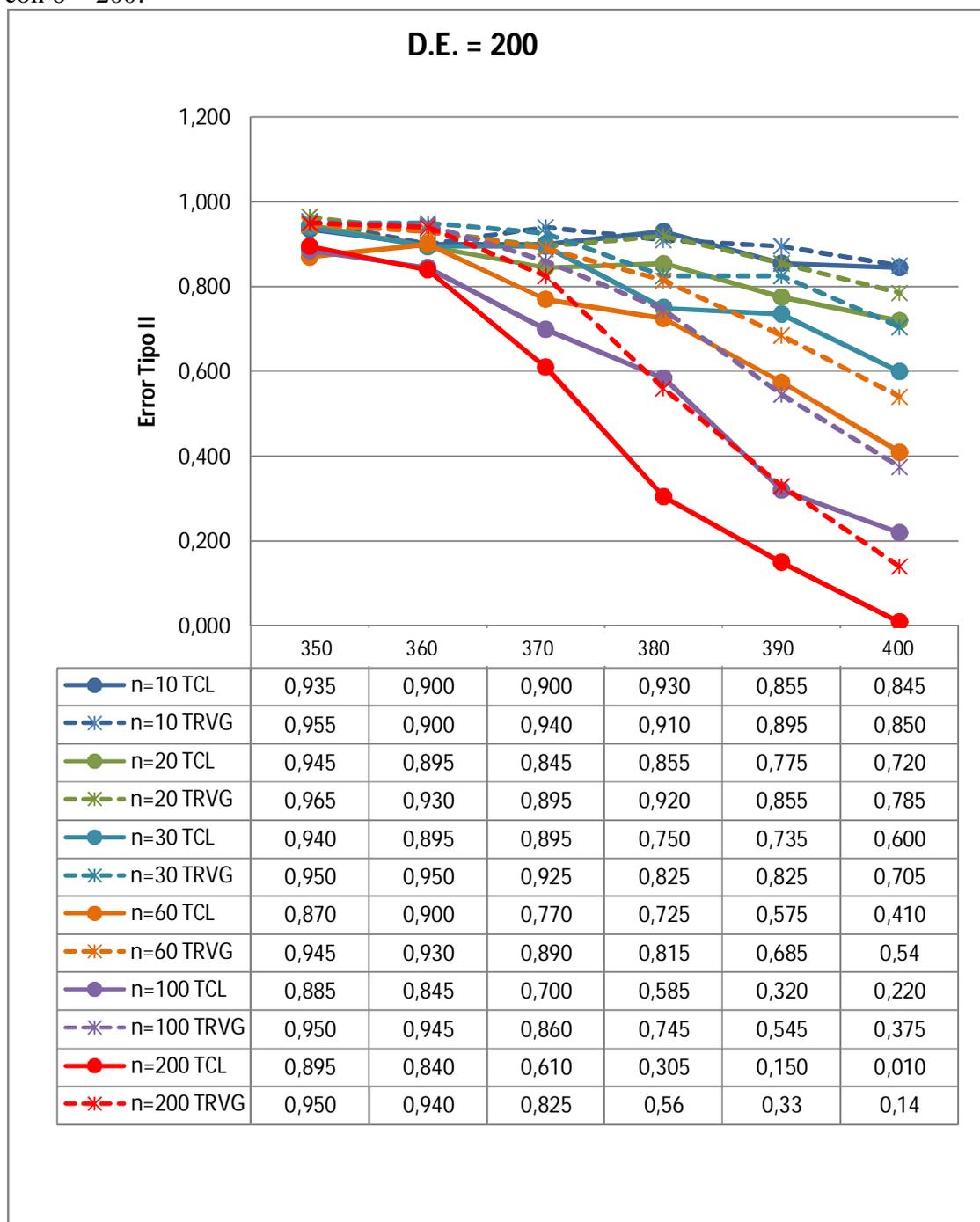
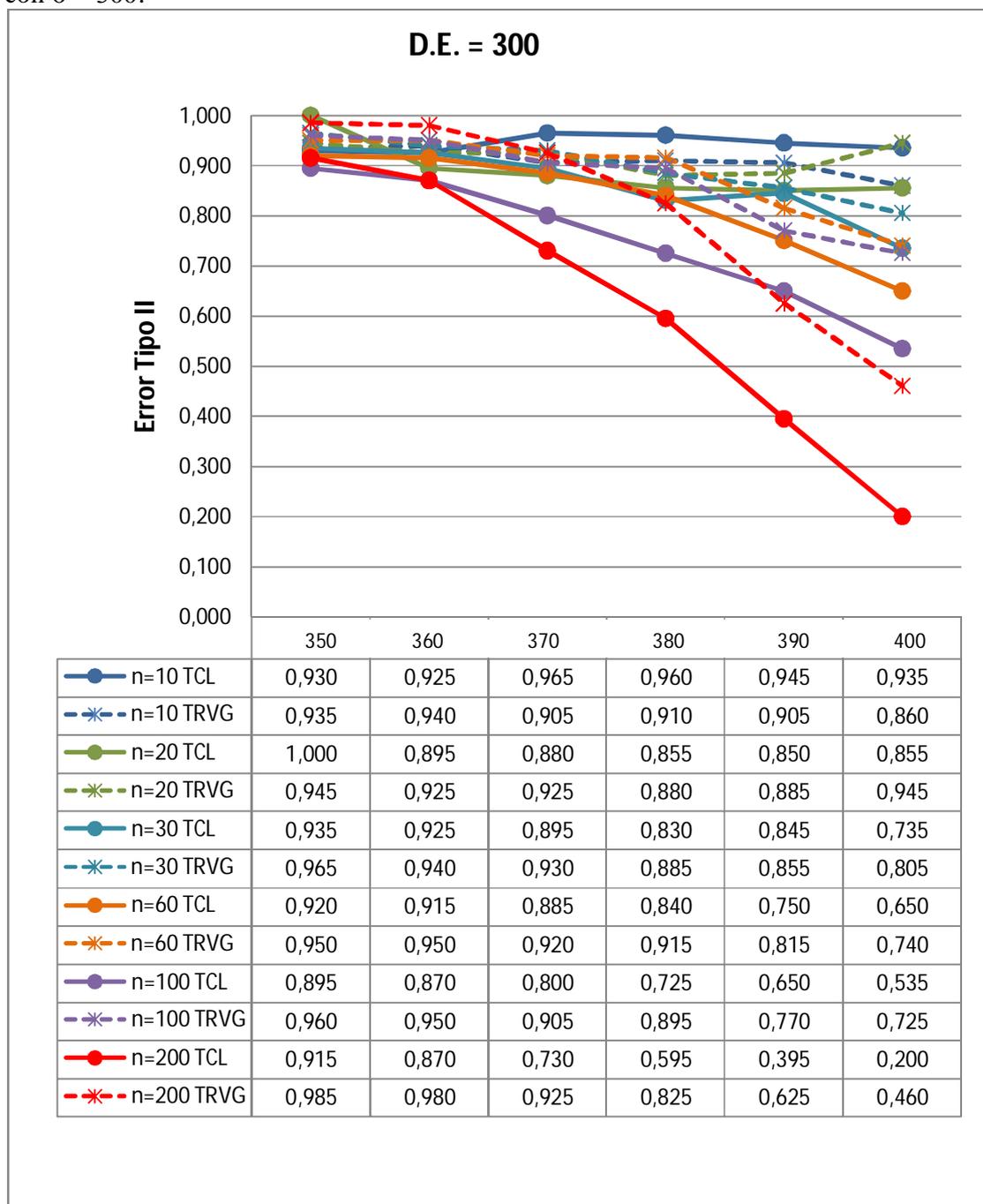


Figura 8. Comportamiento del error tipo II, según media poblacional y tamaño muestral, con  $\sigma = 300$ .



#### 4. CONCLUSIÓN

Para analizar la convergencia de la aproximación Chi-cuadrado, se debe evaluar en función de los tamaños muestrales, puesto que su comportamiento depende de éste. En efecto, para tamaños muestrales pequeños (menores a 60), el error tipo I es menor al emplear el test T, independiente de la desviación estándar; mientras que para tamaños muestrales mayores, se comporta mejor la aproximación Chi-cuadrado. En cuanto al error tipo II, sea cual sea el tamaño muestral y la desviación estándar, el test T parece comportarse siempre será mejor; esto es, proporcionando una mejor protección contra el error Tipo II.

Cabe notar además cómo la combinación de tamaños muestrales superiores a 100 y desviaciones estándar mayores a 200, hacen que ambos estadísticos tiendan a comportarse de manera similar y en torno a un error tipo I de 0,05 (fig. 3 y 4)

Otro aspecto importante a ser destacado, es cómo ambos estadísticos de prueba minimizan rápidamente su error tipo II cuando se tienen desviaciones estándar bastante pequeñas y grandes tamaños de muestra (fig. 5). Lo que resulta totalmente opuesto cuando las varianzas son demasiado grandes (fig. 8)

## 5. REFERENCIAS

BARTOSZYŃSKI, R; NIEWIADOMSKA – BUGAJ, M. (2008). *Probabiliy and Statistical Inference*. Wiley. New York. **13**, 491 – 494.

KENDALL, M.; STUART, A. (1960); *The Advanced Theory of Statistics, Volumen 2*. Hafner Publishing Company. New York. **24**, 224 – 227.