

VALORACIÓN DE LA ROBUSTEZ DE LA PRUEBA T DE FRENTE A LA HETEROGENEIDAD DE
VARIANZA Y LA COMPARACION CON LOS METODOS BOOTSTRAP Y LAS PRUEBAS
ALEATORIZADAS

SANDRA PATRICIA ESPINOSA CASTRO

YAMILEC URRUTIA BUENAÑOS.

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERIAS
ESCUELA DE INGENIERIA INDUSTRIAL Y ESTADISTICA
PROGRAMA ACADÉMICO DE ESTADÍSTICA
SANTIAGO DE CALI 2012

VALORACIÓN DE LA ROBUSTEZ DE LA PRUEBA T DE FRENTE A LA HETEROGENEIDAD DE
VARIANZA Y LA COMPARACION CON LOS METODOS BOOTSTRAP Y LAS PRUEBAS
ALEATORIZADAS

SANDRA PATRICIA ESPINOSA CASTRO

YAMILEC URRUTIA BUENAÑOS.

Trabajo de grado presentado para optar al título de estadístico

Director:

ROBERTO BEHAR GUTIÉRREZ.

Codirector

RAFAEL KLINGER ANGARITA

UNIVERSIDAD DEL VALLE

FACULTAD DE INGENIERIAS

ESCUELA DE INGENIERIA INDUSTRIAL Y ESTADISTICA

PROGRAMA ACADÉMICO DE ESTADÍSTICA

SANTIAGO DE CALI, 2012

Agradecimientos

Hoy culmino una gran etapa en mi vida y quiero primero que todo agradecerle a Dios por estar siempre a mi lado dándome su amor incondicional por que sin el nada hubiera sido posible, le agradezco a mi madre y padre Eduarda Buenaños y Alejandro Urrutia por su confianza y amor, a mis maestros, amigos y familiares que estuvieron con migo en este camino tan lindo y lleno de tantas pruebas que pienso que sin su apoyo esto no hubiera sido posible, para terminar le agradezco mucho también a los maestros Roberto Behar, Rafael Klinger y a mi amiga Sandra por estar en esta lucha.

Amor y fe en Dios.

Yamilec

A Dios, mi esposo, mi hijo, familiares y amigos.

Sandra

Tabla de Contenido

1.	Introducción	1
2.	Planteamiento del Problema	3
3.	Objetivos.....	4
3.1.	Objetivo General.....	4
3.2.	Objetivos Específicos.....	4
4.	Justificación.....	5
5.	Antecedentes	6
6.	Marco Teórico.....	8
6.1.	Prueba T Student	8
6.2.	Pruebas Aleatorizadas.....	9
6.2.1	Determinación del número de permutaciones necesaria.....	18
6.3.	Bootstrap.....	13
6.3.1	Técnicas Clásicas de Estimación Bootstrap.....	13
7.	Metodología.....	17
7.1	Generación de Datos Bajo la Distribución Normal.....	19
7.2	Indicador de Robustez.....	20.
7.3	Evaluación de la Robustez de la Prueba T.....	21
8.	Resultados.....	23
9.	Conclusiones y Recomendaciones.....	30
10	Referencias Bibliográficas.....	31
11	Anexos.....	33

Resumen

En este trabajo se evaluará la robustez de la prueba t de student, mediante la comparación de los métodos bootstrap y las pruebas aleatorizadas; observando su comportamiento para diferentes valores de las varianzas en dos poblaciones en términos de la significancia estadística, a cada una de ellas se le simularán unas situaciones en las cuales habrá diferentes parámetros de variabilidad y a la vez un mismo parámetro de centramiento, el análisis se realizará a través del software estadístico R.2.9.

Palabras Clave: Bootstrap, prueba t, nivel de significancia, pruebas aleatorizadas.

1. Introducción

El objetivo de la presente investigación es evaluar la robustez de la prueba t y compararla con las pruebas aleatorizadas y el bootstrap, mediante el método de simulación Monte Carlo. Para estudiar la eficiencia de estos procedimientos en cuanto a la significancia es decir el error tipo I.

Se manipularon tres variables que son: Tamaño de la muestra, varianza de las poblaciones, relación entre el tamaño de las poblaciones y las varianzas de estas.

La prueba t fue inventada por William Sealey Gosset en 1908, para probar igualdad de poblaciones en muestras pequeñas con distribución normal y varianza constante, esta fue desarrollada y generalizada más adelante por Fisher para más de dos medias y se le denomina análisis de varianza.

Esta técnica denominada análisis de varianza es un método utilizado en la actualidad por muchas personas pero hay que tener precaución al utilizarla ya que si no se conoce como funciona esta clase de prueba, se puede llegar a conclusiones erróneas e inútiles para el estudio en el que se esté aplicando esto.

Lo que concierne al método estadístico para la comparación de medias (prueba t-student) cuando se realiza o se aplica el método de comparación de medias con varianzas desconocidas no se refieren por lo general a la distribución del siguiente estadístico y como se ve afectada con cambios en la magnitud de las varianzas. (Montgomery, 2004):

Del estadístico t se dice que “la distribución de la estadística se desconoce, pero se aproxima a una distribución t-student donde los grados de libertad se calculan de acuerdo a las recomendaciones de Gump utilizando la aproximación de Welch - satterthwaite” (Villa Diharce, 2004), parece ser la solución más reportada hasta el momento en algunos de los libros revisados de estadística, pero se observa que son muy pocos los libros los cuales hablan de la distribución y los grados de libertad de esta prueba en especial.

En la presente investigación se pretende analizar la robustez de esta prueba con respecto a la violación del supuesto de igualdad de varianza y su magnitud sobre todo cuando las muestras no son muy grandes, es decir hasta cuanta diferencia en varianzas estas pruebas soportan, para que sigan siendo buenas para la toma de decisiones acerca de si dos o más poblaciones poseen la misma media.

Se valoran los métodos de comparación de medias como el de la t-student, algunos métodos no para métricos como el bootstrap y las pruebas de aleatorización, con respecto a su robustez ante la presencia de violación del supuesto de homogeneidad de varianzas.

La evaluación de la robustez de las diferentes pruebas se hará por medio de simulación para diferentes valores de la varianza, generando datos con distribuciones previamente establecidas.

Esta prueba se ha venido utilizado de manera indiscriminada sin tener precaución del cumplimiento del supuesto de homogeneidad de varianzas; no siempre las poblaciones cumplen con este supuesto; se debe tener precaución a fin de no cometer errores en la toma de decisiones en las conclusiones se podría aumentar la probabilidad de cometer error tipo I por que los estimadores pierden sus propiedades.

Las técnicas no paramétricas conllevan a la pérdida de potencia, en general, frente a las paramétricas. Este inconveniente, en cualquier caso, no justifica la elección indiscriminada de técnicas paramétricas para realizar análisis estadísticos, pues la violación de los supuestos por éstas requeridos incide sobre la estimación de la probabilidad asociada al estadístico, fundamento de la decisión estadística. La pérdida de potencia y la infraestimación de la probabilidad de error tipo I constituyen aspectos fundamentales en la aplicación de la estadística. (Solanas, 1992)

2. Planteamiento del Problema

Encontrar cual es el tamaño de diferencia de varianza para la cual la prueba t sigue siendo buena es decir robusta; teniendo en cuenta que en la realidad las varianzas podrían no ser exactamente iguales, lo que nos lleva a plantearnos: ¿En qué grado es robusta esta prueba a la heterogeneidad de las varianzas?

Algunos textos usados en nuestro medio afirman de manera superficial que cuando los datos no provienen de una distribución con varianzas iguales hay que realizar la prueba con un estadístico el cual tiene una distribución t-student aproximada y no hacen referencia al grado de aproximación asociado a un nivel dado de desigualdad de las varianzas.

En la estadística es importante el cumplimiento de algunos supuestos como por ejemplo la igualdad de varianzas; en especial en este proyecto es interesante observar la eficiencia de la prueba t-student, pruebas aleatorizadas, y el método del bootstrap para analizar que tan robustas son estas pruebas ante la violación del supuesto de homogeneidad de varianzas.

3. Objetivos

3.1. Objetivo General

Valorar la robustez de los procedimientos estadísticos a través de la prueba t-student, la prueba no Paramétrica bootstrap, y pruebas aleatorizadas frente a la violación del supuesto de homogeneidad de la varianza mediante su nivel de significancia, en poblaciones normales.

3.2. Objetivos Específicos

- Valorar la robustez de la aproximación de Welch ante la no homogeneidad de varianza cuando se quiere comparar dos medias con la prueba t-student.
- Comparar la prueba t de student con pruebas alternativas no paramétricas como el bootstrap y las pruebas aleatorizadas en cuanto a la robustez a través del nivel de significancia. usando métodos de computación.

4. Justificación

La validez o cumplimiento del supuesto de homogeneidad de varianza en las pruebas de comparación de medias es importante garantizarlo; es muy conveniente de valorar puesto que permite observar hasta cuanta diferencia o magnitud de diferencia de varianza permite la prueba para su buen funcionamiento, cuando se dice esto nos referimos a la capacidad de esta prueba para rechazar una hipótesis nula falsa.

Esta clase de investigaciones ayudan a resolver problemas como el comportamiento de las diferentes pruebas de homogeneidad de varianza, pero más que eso nos enseña a buscar la prueba más indicada para la investigación que se desee realizar; esta clase de trabajo sirve como apoyo para investigaciones futuras relacionadas con estas clases de pruebas de análisis estadístico.

La elección de la técnica o herramienta estadística apropiada para el análisis de resultados de experimentos e investigaciones es fundamental para darle soporte y validez a las conclusiones e inferencias obtenidas de éste ya que la elección equivocada con lleva a conclusiones no verdaderas sobre la realidad del fenómeno analizado.

Muchas personas facilitando la ayuda del software no validan los supuestos en esta investigación se valoraran los riesgos de esta omisión.

5. Antecedentes

Los siguientes autores se refieren a la prueba t de manera general (Montgomery, 2004), (Mendenhall, 1997), (Romero Villafranca & Ramaje, 2005) en donde se dice que cuando las varianzas son distintas se debe utilizar la modificación de la prueba t además de que el desempeño de ella se verá afectado; otros autores como por ejemplo *Pagano M.* (2006) dice que para la aplicación De la prueba de comparación de dos medias provenientes de dos muestras independientes hay que utilizar la prueba t si las varianzas de las poblaciones son iguales para verificar esto muchos piensan que deberían realizarse alguna prueba preliminar de igualdad de varianzas por ejemplo la prueba F con el fin de decidir cual caso de la prueba t utilizar si la modificada o la original otros dicen que estas pruebas son muy sensibles a la falta de normalidad y por ultimo pagano no recomienda la previa utilización de una prueba de varianzas ya que según este texto ya se ha demostrado que la t modificada resulta mejor cuando las varianzas son iguales.

Este estadístico según (Montgomery, 2004) no se distribuye como t. No obstante, t es una buena aproximación de la distribución de t_0 si se usa “v” para los grados de libertad. Una indicación clara de la desigualdad de las varianzas en una grafica de probabilidad normal sería una situación que requeriría esta versión de la prueba t. El lector no deberá encontrar problemas para desarrollar una ecuación para encontrar ese intervalo de confianza para la diferencia en las medias en el caso de varianzas desiguales.

Como alternativa en la literatura como por ejemplo en (Siegel, 1970) se han propuestos pruebas menos robustas ante las exigencias de supuestos como la prueba t; de las cuales a continuación se mencionaran algunas pruebas no paramétricas para comparar dos y más de dos poblaciones independientes como las pruebas, χ^2 , la U de Mann Whitney, Probabilidad exacta de Fisher, prueba de dos muestras de Kolmogorov Smirnov.

En (Welch, the significance of the difference between two means when the population variances are unequal, 1938) Como se supone que la varianza es constante se puede decir que la igualdad de varianza está siendo considerada como parte de la prueba ya que si esto no se cumple no hay forma de contrastar la igualdad de medias, además en el criterio $u =$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma_1 + \Sigma_2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$
 las diferencias de las desviaciones pueden enmascarar las diferencias de las

medias con lo que se llega a la conclusión que las medias no sean significativamente diferente en la mayoría de los casos cuando deberían serlo.

El criterio $v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma_1}{n_1(n_1-1)} + \frac{\Sigma_2}{n_2(n_2-1)}}}$ su distribución depende del cociente de varianzas $\theta = \frac{\sigma_1^2}{\sigma_2^2}$.

Para ver esto este autor explica por medio de un ejemplo con el cual llega a conclusiones como la siguiente si $\sigma_1^2 < \sigma_2^2$ la significancia de la prueba u se subestima y si ocurre lo contrario se sobreestima, cuando $n_1 = n_2$ los criterios u y v son iguales.

El criterio u no se distribuye t cuando las varianzas son desiguales y sus resultados no son los mejores, en el ejemplo tienen 20 datos y tres escenarios que son los siguientes: (a). $n_1 = n_2 = 10$, $p(|u| > t_{1,0.025} = 2.101)$; (b) $n_1=5$, $n_2=15$, $p(|u| > t_{1,0.025} = 2.101)$; (c) $n_1=5$, $n_2=15$, $p(|v| > t_{1,0.025} = 2.101)$. donde se encontró que cuando el cociente de varianzas aumenta y el tamaño de muestra es diferente el criterio u se comporta muy diferente a lo esperado con una probabilidad de rechazo del 30%, lo cual indicaría la no robustez ante cambio de tamaño de muestra y diferencia de varianza, el criterio v es menos sensible a este cambio de varianza según esto pero cuando el cociente está entre los valores de 1.01- 1.1 la probabilidad de rechazo esta alrededor del 5% y cuando el cociente tiende a infinito la probabilidad llega alrededor del 10% lo cual indica que v se comporta mejor cuando las varianzas son distintas y tenemos tamaños desiguales.

6. Marco Teórico

6.1. Prueba t Student

Cuando las varianzas son desconocidas R. A. Fisher ha propuesto a una solución en (Welch, the significance of the difference between two means when the population variances are unequal, 1938) la cual consiste en utilizar un estadístico que tiene una distribución aproximada a la t de student donde la hipótesis nula es:

$$H_0 : \mu_1 = \mu_2$$

El estadístico $T_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ (6.1.1) tiene una distribución que es aproximadamente una

distribución t con grados de libertad dados por:

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 + 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 + 1} \left(\frac{s_2^2}{n_2} \right)^2} - 2 \quad (6.1.2)$$

Si la hipótesis nula $H_0 : \mu_1 = \mu_2$ es verdadera. Por tanto si $\sigma_1^2 \neq \sigma_2^2$ la hipótesis de la ecuación (6.1.1) se prueba así. Se calcula el estadístico de prueba T_0^* que tiene una distribución t aproximada con v grados de libertad, entonces si

$$t_0^* > t_{\alpha/2, v} \\ \text{o si} \\ t_0^* < -t_{\alpha/2, v}$$

Debe rechazarse $H_0 : \mu_1 = \mu_2$.

Las alternativas unilaterales se tratan de manera similar. Para probar

$$\begin{aligned}H_0 : \mu_1 &= \mu_2 \\H_1 : \mu_1 &> \mu_2\end{aligned}$$

Se calcula el estadístico de prueba t_0^* $H_0 : \mu_1 = \mu_2$ $t_0^* > t_{\alpha, v}$

Para la otra hipótesis alternativa unilateral,

$$\begin{aligned}H_0 : \mu_1 &= \mu_2 \\H_1 : \mu_1 &< \mu_2\end{aligned}$$

Se calcula el estadístico de prueba t_0^* $H_0 : \mu_1 = \mu_2$ $t_0^* < -t_{\alpha, v}$ ”

6.2. Pruebas Aleatorizadas

Una prueba de asignación al azar está definida como una prueba de permutación basada en la aleatorización (asignación aleatoria) para probar una hipótesis nula sobre los efectos del tratamiento en un experimento aleatorio. La prueba se llevará a cabo de la siguiente manera. Una prueba estadística se calcula para los datos experimentales, los datos se permutan (dividido o reordenar) varias veces y la prueba estadística se calcula para cada una de las permutaciones de los datos resultantes. Estas combinaciones de datos, incluyendo una representación de los resultados obtenidos, constituyen la significancia (Edgington 1980).

Pero veamos el contexto el contexto histórico en el que surgieron las pruebas de aleatorización según Noreen (1989, p. 32).

“Fisher origino el concepto de prueba de aleatorización en su libro “The Desing of the experimENTS” publicado 1935. En las dos décadas que siguieron a la publicación de este libro,

las pruebas de aleatorización atrajeron la atención de estadísticos teóricos como Pitman, Pearson, Scheffe, Noether, Lehmann y Stein, y Hoeffding. La aplicación de las pruebas de aleatorización para la comprobación de hipótesis reales fue impedida, sin embargo, por el alto costo que suponía recalcular manualmente la prueba estadística tantas veces. Los estadísticos acabaron con esta dificultad práctica construyendo tablas de probabilidad para datos genéricos (por ejemplo rangos y categorías nominales). El resultado final de este esfuerzo es el amplio rango de pruebas no paramétricas fundadas en aquellas referencias estándar como Siegel (1956) y Hollander y Wolfe (1973).

Genuinamente, todas las pruebas no paramétricas son casos especiales de pruebas de aleatorización exactas en las cuales las observaciones son, o han sido remplazadas por rangos o categorías nominales.

Las pruebas de aleatorización como “procedimientos para determinar la significación estadística directamente desde los datos experimentales sin recurrir a tablas de significación”. En las pruebas de aleatorización la población de las permutaciones representa el modelo de independencia Edgington (1980, p.1).

La principal ventaja que ofrecen las pruebas de aleatorización es, sin duda, su gran versatilidad y aplicabilidad ya que pueden ser empleadas para hallar la significación tanto de pruebas simples como de pruebas más sofisticadas, incluso cuando las distribuciones muestrales de estadísticos de contraste de dichas pruebas no están tabuladas, o en el caso que no puedan utilizarse por incumplir alguno de los supuestos teóricos en que se basan (muestreo aleatorio, normalidad, homogeneidad de varianza, etc).

El valor del grado de significación obtenido de una prueba de aleatorización es de fácil interpretación: no es más que la interpretación de una proporción observada.

Por otra parte, un aspecto fundamental que caracteriza a las pruebas de aleatorización es su validez para cualquier tipo de muestra, tanto si ha sido extraída aleatoriamente de una población como si no. Esta es una de las propiedades más importantes de estos procedimientos, debido al uso común de muestras no aleatorias en la experimentación y a que las pruebas basadas en las distribuciones clásicas (t, F,...) pueden no ser válidas para este tipo de muestras.

Una de las desventajas que presentan las pruebas de aleatorización es que no permiten, en sentido estricto, realizar inferencias a la población origen de la muestra, aunque ésta haya sido extraída por procedimientos de muestreo aleatorio.

Sobre estos inconvenientes, Noreen (1989) apunta, sin embargo, que si se rechaza una hipótesis nula con base en una prueba de aleatorización aplicada a una muestra realmente representativa de la población es probable que también pudiera rechazarse la hipótesis en dicha población. Es decir que los procedimientos de extracción o elección de los sujetos y/o el diseño de la investigación pueden utilizarse como soportes “casi formales” para realizar inferencias a la población.

6.2.1 Determinación del número de permutaciones necesarias

Las pruebas de aleatorización se pueden aplicar realizando todas las posibles permutaciones de los datos, o a partir de una muestra aleatoria de permutaciones.

Cuando se realizan todas las permutaciones posibles de los datos de unas variables relativas a las otras (permutación sistemática), la prueba se denomina prueba de aleatorización exacta. Sin embargo, en situaciones en las que el número de permutaciones posibles es tan grande que imposibilita la aplicación de la prueba de aleatorización exacta, se realizan permutaciones aleatorias de los datos y, en este caso, la prueba se denomina prueba de aleatorización aproximada (Noreen, 1989).

En las pruebas de aleatorización aproximadas, el procedimiento consiste en calcular el estadístico de interés para el conjunto original de datos y para una muestra aleatoria de permutaciones de estos datos. Esto supone generar un primer conjunto de datos por permutación aleatoria de los datos originales, y a continuación ir generando, también por permutaciones aleatorias, nuevos conjuntos de datos, siempre a partir del último conjunto generado.

Cuando se realizan pruebas de aleatorización aproximadas, es interesante conocer entre qué valores oscilarán las estimaciones realizadas por aleatorización aproximada para un determinado nivel de significación p obtenido mediante la realización de todas las posibles permutaciones (test de aleatorización exacto), y un determinado número de permutaciones NP (Manly, 1991, p.34). el nivel de significación estimado se distribuirá aproximadamente según una ley normal alrededor de p con varianza $p(1-p) / N$ cuando el número de permutaciones aleatorias NP sea grande y es de la forma (6.2):

$$p \pm z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{NP}} \quad (6.2)$$

Por otro lado, el número de permutaciones posibles (NP+1) de los datos se determina con base en las fórmulas de combinatoria presentadas en la tabla siguiente donde n simboliza el número de casos de la muestra, n_i el número de observaciones en cada uno de los k niveles de la variable independiente (k simboliza también el número de variables en el diseño de datos apareados)

Numero de permutaciones en función del diseño de la investigación

Tipo de Diseño	No de Permutaciones Posibles (NP+1)
Datos Independientes	$\frac{n!}{n_1! n_2! \dots n_k!}$
Datos apareados	$(k!)^n$ ó $n!$

Las pruebas de aleatorización

“Prueba de aleatorización” o “randomization test” en la parte algorítmica su procedimiento es el de determinar el grado de significación mediante la permutación de los datos experimentales y el cálculo repetido de una prueba estadística.

1. Elegir la prueba estadística de su objetivo.
2. Calcular $\hat{\theta}_0$ para el conjunto de datos originales X_0 .

3. Inicializar el contador NSIG = 0
4. Fijar el número de permutaciones (NP) que se realizarán con los datos.
5. Para $i = 1$ hasta (NP)
 - 5.1. $X_1^* = \text{PERM}(X_0)$.
 - 5.2. Calcular el estadístico de interés $\hat{\theta}_i^*$ para el conjunto de datos generados X_i^*
 - 5.3. Si $\hat{\theta}_i^* \geq \hat{\theta}_0$ o según sea el caso $\hat{\theta}_i^* \leq \hat{\theta}_0$
 $\text{NSIG} = \text{NSIG} + 1$
6. Calcular el grado de significación. $(\text{NSIG}+1)/((\text{NP}) + 1)$

6.3. Bootstrap

En el año de 1979, y a partir de un profundo y reflexivo análisis sobre el origen y evolución de los métodos estadísticos, tanto a nivel teórico como aplicado, Bradley Efron desarrolla la técnica Bootstrap.

Concretamente, la idea básica sugerida por Efron (1979) es la siguiente:

Si una muestra aleatoria contiene la máxima información disponible sobre la población, ¿por qué no proceder como si la muestra fuese la población y entonces estimar la distribución muestral de un estadístico generando nuevas muestras mediante un muestreo aleatorio con reposición, a partir de los datos de la muestra original?.

Efron describe el bootstrap como una técnica útil para evaluar el error estadístico (error estándar, sesgo y probabilidad de error de predicción) y para realizar estimaciones de parámetros., no obstante rápidamente se describen diversas adaptaciones y desarrollos para obtener el grado de significación en las pruebas de hipótesis”.

6.3.1. Técnicas Clásicas de estimación Bootstrap

En las siguientes secciones de este apartado se mencionaran algunas de las técnicas del bootstrap introducidas por Efron.

a. Bootstrap no Paramétrico

La estimación bootstrap no paramétrico se basa en el supuesto de que la función de distribución empírica $\hat{F}_n$ obtenida a partir de la muestra correspondiente a extracciones aleatorias $\{x_1, x_2, \dots, x_n\}$ de n variables aleatorias independientes X_i con idéntica función de distribución F :

$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$$

Es la estimación máximo verosímil no paramétricas de F , fundamentada en la asignación de probabilidad $1/n$ a todos y cada uno de los datos muestrales (Efron 1979b, 1982^a). En este sentido, Bickel y Freedman(1981), en su estudio sobre las propiedades asintóticas del bootstrap demuestran el postulado anterior a partir del teorema de Glivenko-Cantelli, principio que permite establecer una convergencia casi segura, cuando $n \rightarrow \infty$, entre las distribuciones F y $\hat{F}_n$

$$|F - \hat{F}_n| \rightarrow 0$$

Consecuentemente, se puede considerar que $F = \hat{F}_n$, proposición que sustenta al bootstrap no paramétrico.

Siguiendo a Sánchez (1990a) la forma habitual de estudiar la distribución (asintótica). De un estadístico se estudia en dos etapas:

1. Encontrar una aproximación (large sample approximation) para la distribución muestral cuando $n \rightarrow \infty$
2. Sustituir F por $\hat{F}_n$ en la distribución asintótica

El problema que presenta esta aproximación reside en la complejidad de la distribución asintótica. El bootstrap soluciona este al sustituir F por $\hat{F}_n$ antes de hallar la distribución muestral, la cual es estimada a través del repetido muestreo con reposición a partir de los datos de muestra original.

Considerando lo anterior, el algoritmo general de la estimación bootstrap no paramétrico se basará en la extracción aleatoria y con reposición de un conjunto de muestras X_b^* partir de los datos contenidos en la muestra original X_0 , calculando el estadístico o estadísticos $\hat{\theta}_b^*$ de interés para cada una de las muestras bootstrap (el supraíndice “*” indica que se trata de una muestra extraída de X_0 y el subíndice “b” indica que se trata de la b muestra extraída de X_0).

b. Bootstrap Paramétrico.

En el caso de que la función de distribución F_θ a partir de la cual se han extraídos los datos sea conocida excepto en su parámetro θ , puede recurrirse a la estimación de este parámetro considerando que $\hat{\theta}$ es una buena estimación de θ obtenida a partir de los datos muestrales y por tanto que $F_\theta \approx F_{\hat{\theta}}$.

En este caso en lugar de extraer muestras aleatorias a partir de la muestra original X_0 , pueden extraerse directamente por muestreo Monte Carlo con base en el modelo de distribución $F_{\hat{\theta}}$. Este proceso se corresponde con el denominado bootstrap paramétrico.

El algoritmo general del bootstrap paramétrico es igual al del bootstrap no paramétrico, excepto en el paso en que se extrae la muestra bootstrap X_b^* .

Algoritmo bootstrap paramétrico

- a. Definir el estadístico de interés θ .
- b. Calcular $\hat{\theta}_0$ para la muestra de datos original X_0 .
- c. Fijar el número B de remuestras a efectuar
- d. Para $b = 1$ hasta B.

Extraer una muestra bootstrap X_b^* a partir del modelo de distribución $F_{\hat{\theta}}$:

$X_b^* = \text{DISTRIBUCION}(n, \hat{\theta}_0)$

Calcular el estadístico $\hat{\theta}_b^*$ para los datos de X_b^*

- e. Calcular la estimación bootstrap del parámetro $\hat{\theta}^* = \sum \hat{\theta}_b^* / B$.

7. Metodología

Con el fin de evaluar la robustez de la prueba t, el bootstrap y las pruebas aleatorizadas se llevo a cabo un estudio de simulación, se hizo uso de la distribución normal, en distintos escenarios bajo la variación de las varianzas y una única media igual a cero descritos mas adelante. En ellas se comparó la robustez de las pruebas propuestas cuándo se incumplía el supuesto de varianzas homogénea, para tal fin se manipularon:

1. el tamaño de muestra
2. La relación entre el tamaño de muestra y la varianza
3. Varianzas. Estas variables fueron seleccionadas por ser de uso habitual en otras investigaciones similares a la nuestra como el artículo base de esta investigación (Welch, the significance of the difference between two means when the population variances are unequal, 1938).

El número de repeticiones, es decir el numero de remuestras extraídas, será de 5.000 siguiendo las recomendaciones de Chernick (1999, p. 114), que considera esta cifra adecuada para la mayor parte de los problemas planteados.

El comportamiento de estas de la prueba t, el bootstrap y las pruebas aleatorizadas fueron investigados utilizando $n_1=n_2=10,15, 25, 30, 40,50$ y 100 y las varianzas son tomadas de las siguientes manera $\sigma_{2i}^2 = k\sigma_{1i}^2$ y $\lambda_i = \frac{\sigma_{1i}^2}{n_{1i}} + \frac{\sigma_{2i}^2}{n_{2i}}$ para cada muestra.

Se define que

$$\kappa = \frac{\sigma_2^2}{\sigma_1^2} \quad (7.0.1)$$

y

$$\lambda = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad (7.0.2)$$

Entonces a partir de estas dos ecuaciones que se denominaron parámetros de la simulación se haya el valor de σ_1^2 y σ_2^2 que corresponden al valor de la varianza de cada población., de la siguiente manera:

$\kappa = \frac{\sigma_2^2}{\sigma_1^2}$ Entonces si en ambos lados se multiplica por σ_1^2 esto queda de la siguiente manera.

$$\sigma_1^2 \kappa = \frac{\sigma_2^2}{\sigma_1^2} \sigma_1^2 \quad \sigma_2^2 = \sigma_1^2 \kappa$$

Teniendo ya σ_2^2 este se procede a encontrar el valor de σ_1^2 a partir de $\lambda = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ entonces se reemplaza el valor de σ_2^2 en términos de σ_1^2 en λ y se obtiene el valor de la σ_1^2 :

$$\lambda = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \text{ Entonces como } \sigma_2^2 = \sigma_1^2 \kappa \text{ se reemplazo esto en } \lambda$$

$$\lambda = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2 \kappa}{n_2} \text{ se saca como factor común } \sigma_1^2 \text{ obteniendo así el resultado siguiente}$$

$$\lambda = \sigma_1^2 \left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right) \text{ Multiplicando a ambos lados por } \frac{1}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)} \text{ finalmente se haya el valor de } \sigma_1^2.$$

$$\frac{1}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)} * \lambda = \sigma_1^2 \left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right) * \frac{1}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)}$$

$$\frac{1}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)} * \lambda = \sigma_1^2$$

$$\sigma_1^2 = \frac{\lambda}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)}$$

$$\text{Finalmente se tiene el valor de } \sigma_1^2 = \frac{\lambda}{\left(\frac{1}{n_1} + \frac{\kappa}{n_2} \right)} \text{ y el valor de } \sigma_2^2 = \sigma_1^2 \kappa.$$

Para el análisis la robustez de las pruebas de comparación de medias t-student y análisis de varianza se utilizara el software *r* Project dada su flexibilidad para la creación de algoritmos particulares.

La prueba t se compararan con las pruebas aleatorizadas y el bootstrap bajo distintos escenario que se definirán más adelante para analizar la robustez de estas ante la violación del supuesto de homogeneidad de varianza. La distribución con la cual se trabajara es la normal y sus diferentes parámetros y tamaños de muestra los cuales se especificaran más adelante; lo cual se evaluara mediante el nivel de significancias de ellas; para lograr identificar factores que ayuden a aplicar la prueba en donde realmente sea útil para obtener o tomar decisiones fiables para el cumplimiento de los objetivos.

7.1. Generación de Datos Bajo la Distribución Normal

Se utiliza la función `rnorm()` de la librería `stats` del programa estadístico R 2.12 para generar los datos de cada una de las poblaciones bajo la normalidad con media 0 y varianzas $\sigma_{2i}^2 = k\sigma_{1i}^2$. Entonces:

- Distribución donde provienen los datos

$$X_1 \approx N\left(\mu_1 = 100; \sigma_1^2 = \frac{\lambda}{\frac{k}{n_1} + \frac{1}{n_2}}\right); \quad X_2 \approx N(\mu_2 = 100; \sigma_2^2 = k * \sigma_1^2) \quad (7.1.1)$$

- Se generaran B = 5000 muestras de tamaño $n_i = 10, 15, 25, 30, 40, 50$ y 100. para ambas poblaciones para diferentes cocientes de varianzas

$\kappa = \frac{\sigma_1^2}{\sigma_2^2} = 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, \dots, 2, 2.1, \dots, 3, 3.1, \dots, 4$ y el parámetro

$\lambda = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = 0.1, 0.2, 0.3$ y 0.4 el cual depende de los tamaños de muestras.

Distribución donde provienen los datos.

Población 1 $X_1 \sim N(\mu = 100, \sigma_1^2)$

Población 2 $X_2 \sim N(\mu = 100, k\sigma_1^2)$

λ : es la suma de las varianzas.

K: es el cociente varianza y a medida que este crece aumenta la heterogeneidad.

Donde: se define el cociente de varianzas (k) como el nivel de heterogeneidad es decir es el indicador de la diferencia de una varianza a otra; y la varianza de la diferencia de medias (λ) como indicador de nivel de calidad de estimación.

Escenarios

k	λ	n1	n2	σ^2_2	σ^2_1
1	0,1	5	5	0,250	0,250
1	0,2	5	5	0,500	0,500
1	0,3	5	5	0,750	0,750
1	0,4	5	5	1,000	1,000
,	,	,	,	,	,
,	,	,	,	,	,
,	,	,	,	,	,
2	0,1	5	5	0,333	0,167
2	0,2	5	5	0,667	0,333
2	0,3	5	5	1,000	0,500
2	0,4	5	5	1,333	0,667
,	,	,	,	,	,
,	,	,	,	,	,
,	,	,	,	,	,

7.2 Indicador de Robustez

Para evaluar el desempeño de la prueba t, las pruebas aleatorizadas y el bootstrap se utiliza como indicador el nivel de significancia empírico (PE) el cual es la proporción entre 5000 significancias sobre el número total de significancias menores e iguales que el 5% obtenido en cada una de las 5000 muestras y es de la forma:

$$PE = \frac{\sum_{i=1}^{5000} p_i}{5000} \quad (7.2.1)$$

Donde p_i es # total de significancias menores o iguales a 0.05.

Para cada cambio en la varianza se realizan 1000 simulaciones, teniendo 1000 PE estimadas, a las cuales se haya el valor esperado, obteniendo la significancia estimada por cada cambio tanto en varianza como en tamaño de muestras.

El valor esperado esta dado por:

$$E(PE) = \frac{\sum_{i=1}^{1000} PE_i}{1000} \quad (7.2.2)$$

Entonces para indicar que metodología se utiliza vamos a denominar $E(PE)$ como P_m donde m indica la metodología utilizada, tomando valores 1, 2,3.

1 significa prueba t

2 significa bootstrap

3. significa prueba aleatorizadas.

7.3. Evaluación de la Robustez de la Prueba T.

La robustez de la prueba t se evaluara mediante el la significancia estadística denominada $P_m = E(PE) =$ p-valué calculado.

La hipótesis de la prueba t. es:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

La hipótesis de trabajo es:

Se desea investigar que tan robusta es la prueba t modificada de Welch, ante la no homogeneidad de varianzas para muestras independientes-la prueba t modificada se describe a continuación.

De acuerdo con Welch (1938), El estadístico de prueba utilizado $T_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ tiene una

distribución que es aproximadamente una distribución t con grados de libertad dados por

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 + 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 + 1} \left(\frac{s_2^2}{n_2} \right)^2} - 2$$

Esta es la prueba en la cual se utiliza para la evaluación de las técnicas estadísticas mencionadas anteriormente.

METODO	SIGNIFICANCIA EMPIRICA
PRUEBA T	$P_1 = \frac{\sum_{i=1}^{1000} PE_{1i}}{1000}$
BOOTSTRAP	$P_2 = \frac{\sum_{i=1}^{1000} PE_{2i}}{1000}$
PRUEBA ALEATORIZADAS	$P_3 = \frac{\sum_{i=1}^{1000} PE_{3i}}{1000}$

En el cálculo de la significancia empírica para la prueba t, pruebas aleatorizadas y el bootstrap, se definen tres contadores que son p1, p2y p3 los cuales resumen en su valor el conteo en el cada vez la metodología analizada respectivamente el valor p calculado es menor que el valor p critico, resumido en la tabla anterior.

Si esto es menor que el $\alpha_{teorico}$ de 0.05se concluye, que la prueba es robusta según el método.

8. Resultados

Resultados de la robustez de las diferentes técnicas estadísticas de comparación de medias ante la no homogeneidad de la varianza.

Se presentan a continuación los resultados obtenidos, para evaluar la robustez de la prueba t, se evaluaron 3 tamaños de muestras (10, 15, 25) con pruebas t student, aleatorizadas y bootstrap.

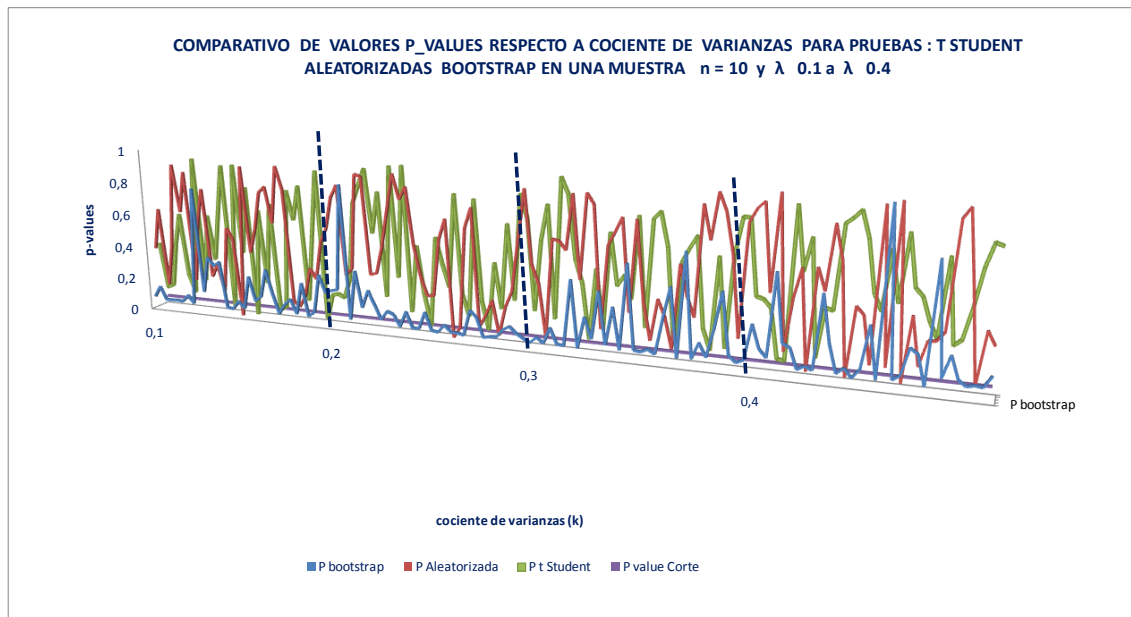


Figura 1

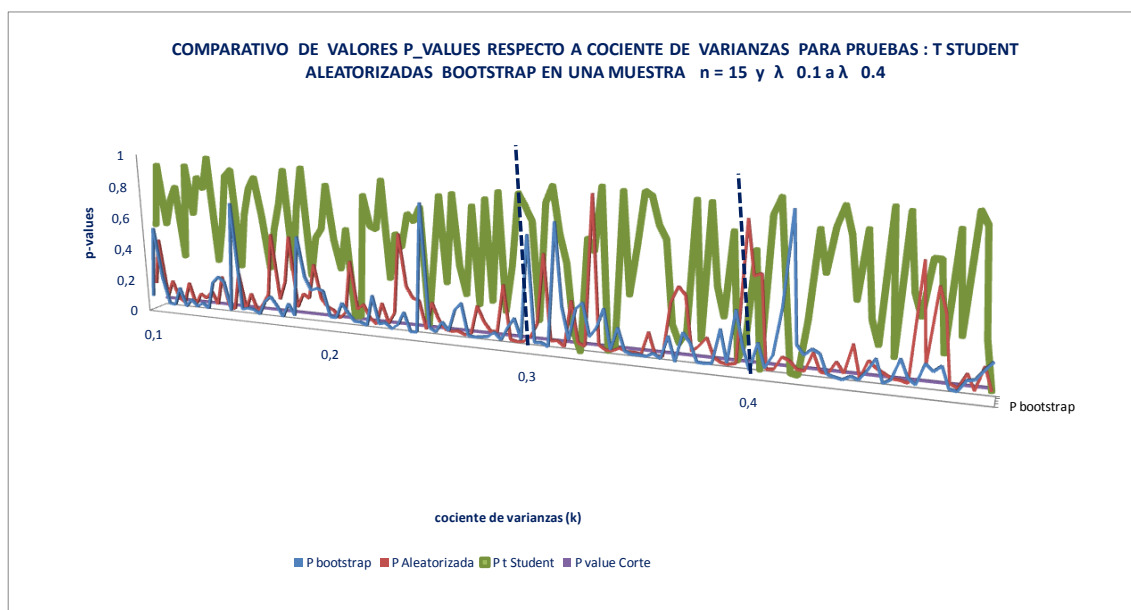


Figura 2

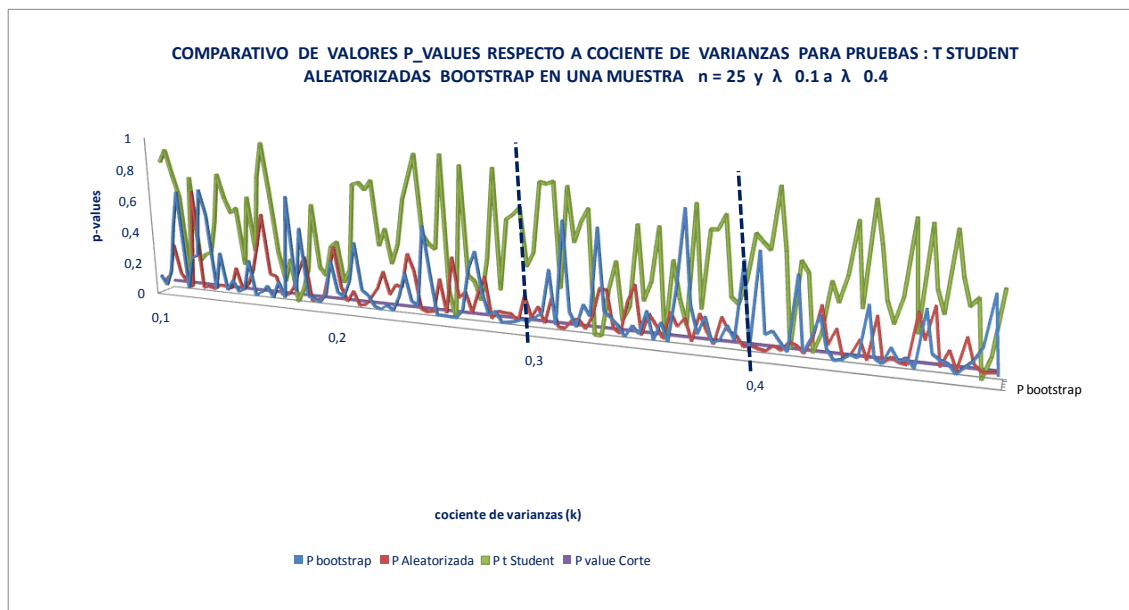


Figura 3

En las figuras 1 a 3, se trazó una línea de referencia en el nivel de significancia deseado ($\alpha = 0.05$), debido a que el objetivo es observar como es el comportamiento de las pruebas utilizadas Cuando los tamaños de muestras varían.

Según los resultados de las figuras 1 al 3, se observa que el mejor caso se da para el método del bootstrap y las pruebas aleatorizadas, puesto que los p-valúes son menores que el valor t critico con respecto a las demás metodología además sus valores tienen un mayor porcentaje de p-valúes por debajo de 0.05 lo cual indica que esta es la mejor en el casos de n igual a 10, en el caso de n igual a 15 es mejor el bootstrap y por ultimo para el caso de n igual a 25 el mejor es la prueba t student.

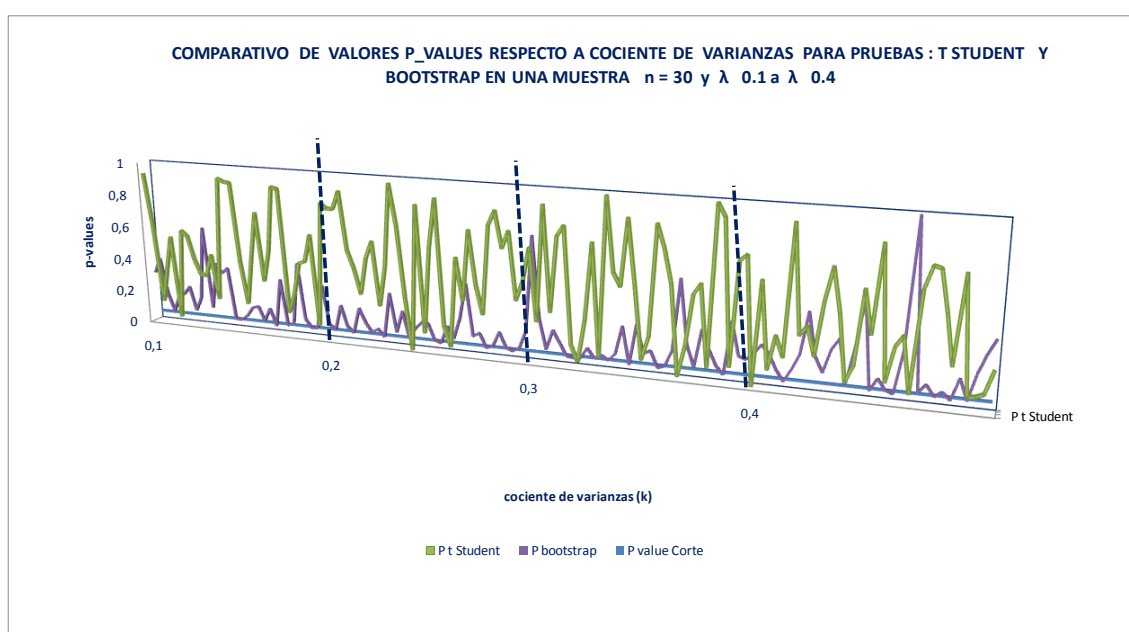


Figura 4

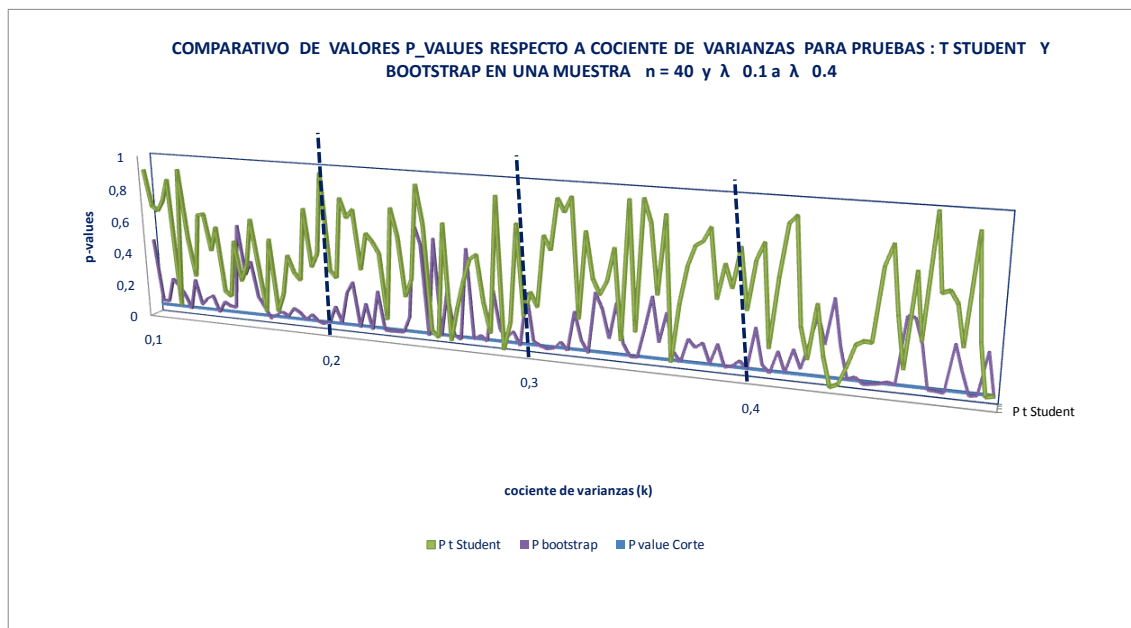


Figura 5

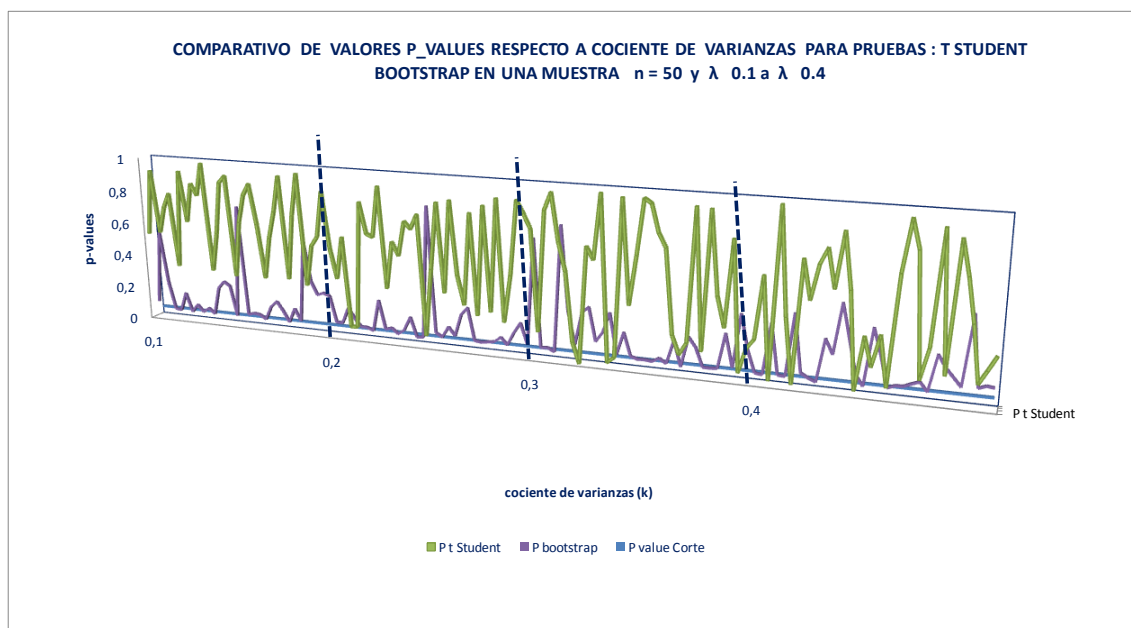


Figura 6

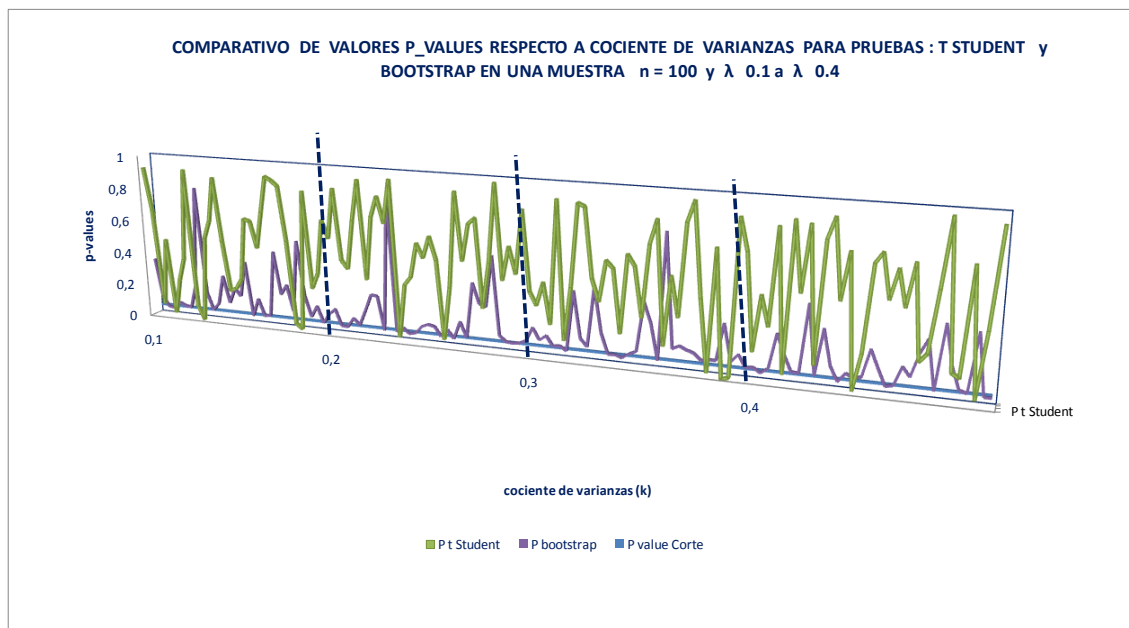


Figura 7

En las figuras 5 a 8 el mejor desempeño es mostrado por método del bootstrap para los tamaños de muestra n igual a 30, 40 y 50 a demás en cuanto al tamaño de muestra crece es mejor la robustez del bootstrap, pero cuando ya el tamaño de muestra pasa a 100 el mejor desempeño lo tiene la prueba t de student.

Cabe también resaltar que cuando el nivel de la calidad de la estimación es 0.4, es en donde la prueba t es más robusta para los tamaños de muestra 30, 40 y 50 y para el método del bootstrap sucede cuando el nivel de la calidad de la estimación es 0.2 ó 0.3 en los tamaño de muestra 30, 40 y 100.

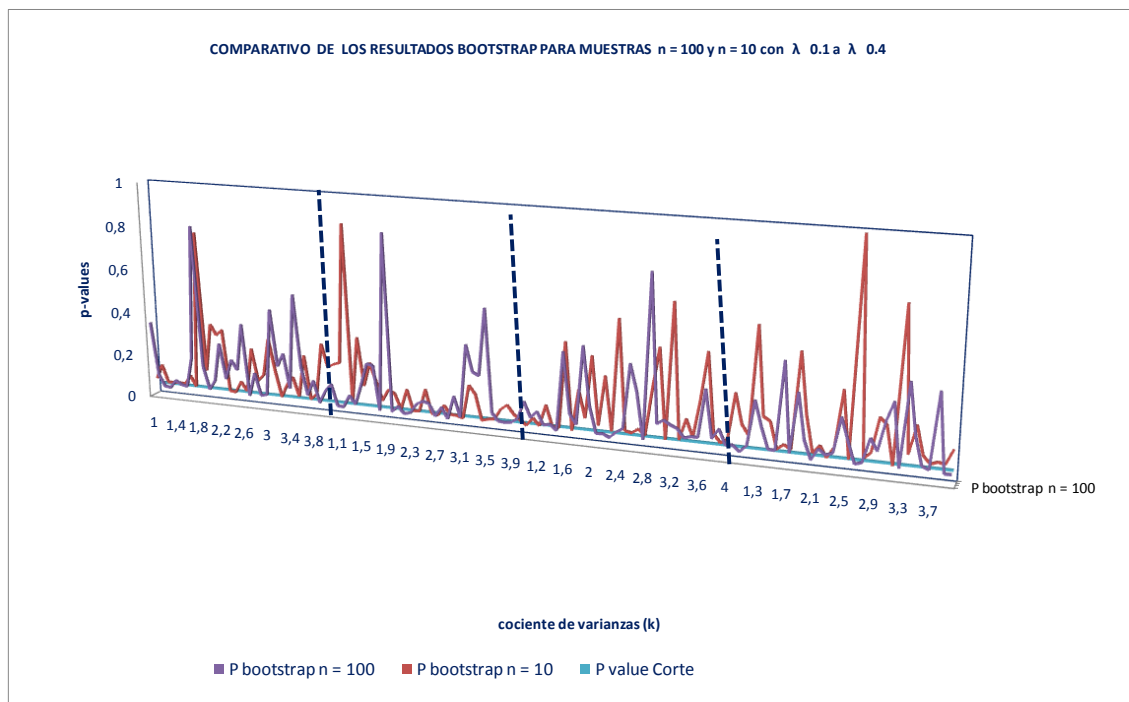


Figura 8.

Se aprecia en la figura 8 que cuando el tamaño de muestra es 100 es más robusto el bootstrap por que el porcentajes de p que están por debajo del 0.05 es mayor que para n igual a 10, lo cual indica que en este caso es mejor cuando se tiene muestras grandes funciona mejor este método para este caso.

También cuando lambda va creciendo a partir de 0.2 la precisión va desmejorándose encontrándose mejor desempeño del método cuando λ es igual 0.2.

Además a partir de los demás tamaños (10, 40 y 100) de muestras se encontró que en la mayoría de los casos el mejor desempeño del método bootstrap se obtiene cuando el nivel de calidad de la estimación (λ) es 0.2 exceptuando para n igual a 50 y n igual 30.

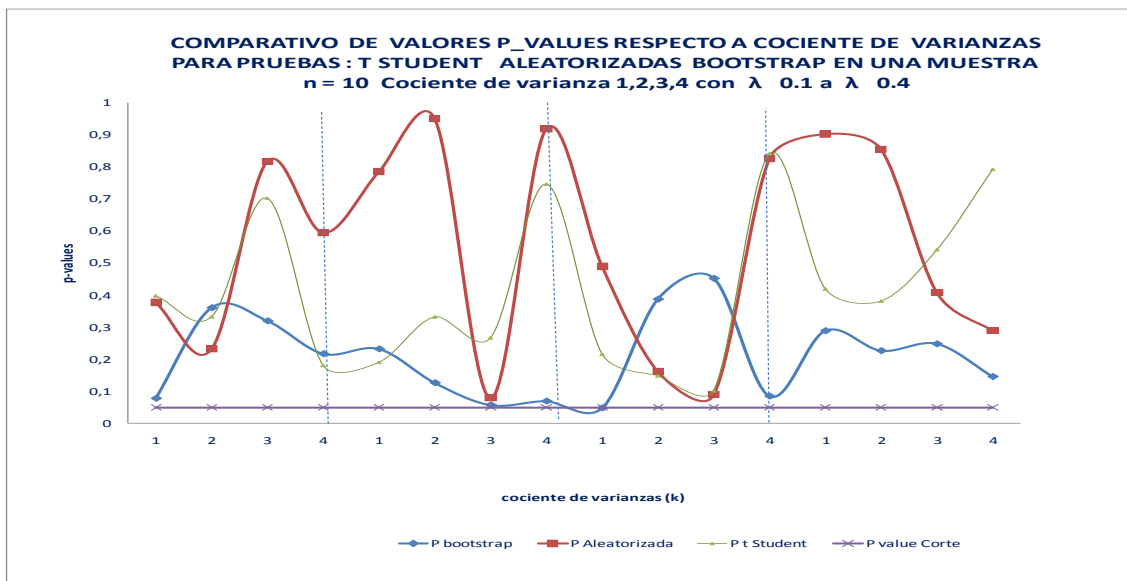


Figura 9

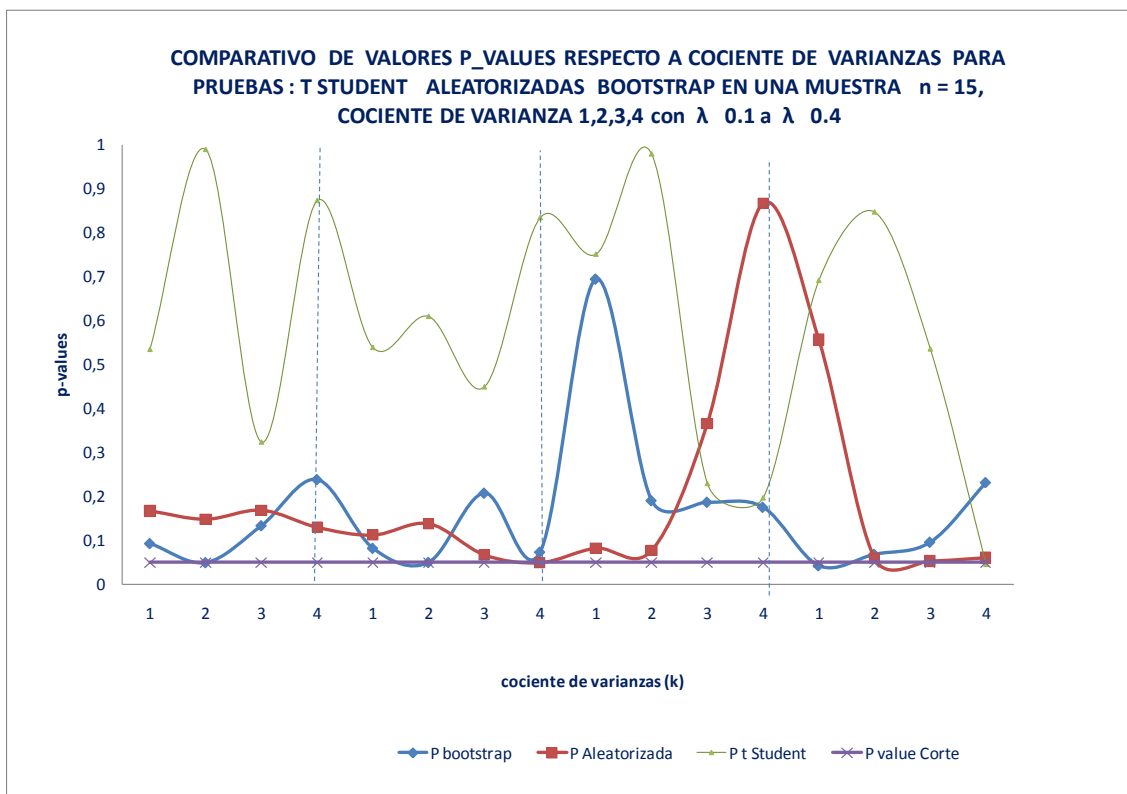


Figura 10

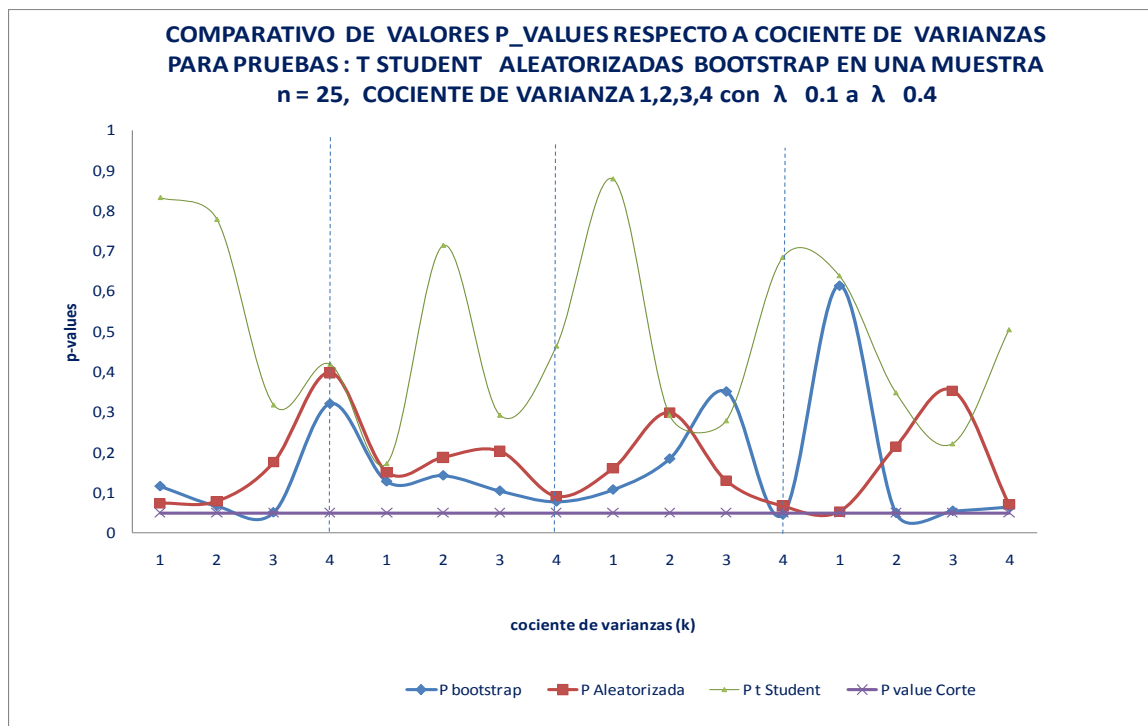


Figura 11

De las graficas 9 a la 11 se observa que el peor desempeño lo muestran en las pruebas aleatorizadas y la prueba t; el bootstrap sobresale con mejores resultados, en el bootstrap cuando k es igual 1 para los tamaños de muestra 10, 15, y para el tamaño de muestra 25. Es robusta cuando k es igual a 2.

Se aprecia también cuando el tamaño de muestra y el valor de k aumenta mejora la significancia para los valores p-valúes en los dos métodos; bootstrap y pruebas aleatorizadas; el bootstrap muestra mejores resultados respecto a las prueba aleatorias puesto que esta presentan p-valúe más pequeños; es decir el bootstrap es mejor que las prueba aleatorias, también cabe notar que cuando el tamaño de muestra es 10 se da la situación que los valores son mayores comparados entre los escenarios.

9. Conclusiones y Recomendaciones

En primer lugar, la evidencia disponible de nuestros experimentos de simulación, no permiten identificar uno de los métodos como uniformemente mejor que los restantes, aun para escenarios particulares definidos por las condiciones del diseño, los niveles de significancia que muestran las probabilidades de rechazar una hipótesis cuando esta es verdadera; en general son lejanas al nivel de significancia 5% especificado en el experimento de simulación.

El método del bootstrap es mejor que el método de las pruebas aleatorizadas y la prueba t de acuerdo con los niveles de heterogeneidad (k) utilizados en esta investigación, es decir cuando las varianzas son iguales o la una es, a lo más, dos veces la otra, esta prueba se comporta mejor que las otras en **cuanto a la robustez**.

La aplicación de un contraste basado en la estadística t de Student modificada de Welch requeriría la comprobación de supuestos previos tales como la normalidad de las variables estudiadas, supuesto difícil de rechazar en un contraste de hipótesis distribucional, en muestras pequeñas. Cuando la distribución del estadístico de contraste no sigue el modelo de la t, bien porque la variable estudiada no se distribuye normalmente en la población o bien porque el tamaño muestral resulta insuficiente, podríamos estar cometiendo un error en la inferencia.

Los resultados arrojados por la investigación no permiten obtener conclusiones contundentes sobre la robustez, pues en varios de los escenarios, se observan comportamientos un poco erráticos, y en ocasiones incoherentes en lo que respecta al impacto del aumento de la heterogeneidad en la calidad de la estimación del nivel de significancia. Muy posiblemente se requiera un mayor rango para el indicador k de heterogeneidad, pues el máximo valor estudiado fue $k=4$, es decir una de las varianzas es cuatro veces la otra.

Por otro lado, es claro en los reportes de la literatura que la efectividad tanto de las estimaciones Bootstrap, como las de las pruebas aleatorizadas, depende en gran medida del acercamiento de la Función de Distribución empírica a la Función de Distribución Poblacional (o de Probabilidad), lo cual se garantiza asintóticamente. Esto implica que dicha cercanía es dudosa en muestras pequeñas. Se recomienda para futuras investigaciones ampliar los niveles de tamaño de muestra a valores mayores.

Sin embargo, aunque los valores no están muy próximos al valor 5% para el nivel de significancia especificado, al comparar los p-valúes en cada método, puede decirse que como tendencia, el método Bootstrap tiene un desempeño más estable.

10. Referencias Bibliográficas.

- [1] Bickel, P. y. (1981). *Some Asymptotic Theory for the Bootstrap*. Recuperado el 4 de 5 de 2011, de http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.aos/1176345637
- [2] Chernick, M. (1999). En *Bootstrap methods: a practitioner's guide*. New York.
- [3] Edgington, E. (1980). En *Randomization Tests* (1972-1991 ed.). New York: Founding.
- [4] Efron, B. (1979b). *Bootstrap methods Another look at the jackknife*. *The annals of Statistics*, 7, 1-26.
- [5] Efron, B. (1982a). *Nonparametric Estimates of Standard Error: The jackknife, the bootstrap and other methods*, *Bometrika*, 68, 589-599
- [6] Hollander, M. y. (1973). En *Nonparametric Statistical Methods*. New York.
- [7] Losilla Vidal, J. M. (s.f.). Recuperado el 25 de Junio de 2010, de <http://www.tesisenred.net/TDX-0701109-141431>
- [8] Manly, B. F. (1991). En *Randomization and Monte Carlo Methods in Biology*. Londres.
- [9] Mendenhall, W. (1997). *Probabilidad y estadística para ingenieros y ciencias*. Mexico: Universidad de Florida.
- [10] Montgomery, D. (2004). *Introducción al análisis de regresión lineal* (Tercera edición ed.). Ceca.
- [12] Moreno, W. (1993). *Aplicaciones al diseño y análisis de experimentos*. Bucaramanga: UIS.
- [13] Noreen, E. (1989). En *Computerintensive methods for testing hypotheses*. New York.
- [14] Pagano Robert, R. (2006). *Estadística para las ciencias del comportamiento* (séptima edición ed.). Thomson.

[15] Romero Villafranca, R., & Ramaje, L. (2005). *Metodos estadisticos en ingenieria*. Universidad politecnica de valencia.

[16] Siegel, S. (1970). *Experimental no parametrico aplicado a las ciencias de la conducta*. Mexico.

[17] Solanas, A. (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de Psicología* , 55, 143-154.

[18] Villa Diharce, E. (2004). El numero efectivo de grados de libertad. *Simposio de petrologia* , 1-7.

[19] Welch, B. L. (Febrero de 1938). *The significance of the difference between two means whwn the population variances are unecual*. *Biometrika Trust* , 350-362
Recuperado el 1 de 7 de 2009, de www.jstor.org/stable/2332010

11. Anexos.

ANEXO I. Escenarios para dos tamaños de muestras 10 y 15.

k	λ	n1	n2	σ_2^2	σ_1^2
1	0,1	10	10	0,500	0,500
1	0,2	10	10	1,000	1,000
1	0,3	10	10	1,500	1,500
1	0,4	10	10	2,000	2,000
1,1	0,1	10	10	0,524	0,476
1,1	0,2	10	10	1,048	0,952
1,1	0,3	10	10	1,571	1,429
1,1	0,4	10	10	2,095	1,905
1,2	0,1	10	10	0,545	0,455
1,2	0,2	10	10	1,091	0,909
1,2	0,3	10	10	1,636	1,364
1,2	0,4	10	10	2,182	1,818
1,3	0,1	10	10	0,565	0,435
1,3	0,2	10	10	1,130	0,870
1,3	0,3	10	10	1,696	1,304
1,3	0,4	10	10	2,261	1,739
1,4	0,1	10	10	0,583	0,417
1,4	0,2	10	10	1,167	0,833
1,4	0,3	10	10	1,750	1,250
1,4	0,4	10	10	2,333	1,667
1,5	0,1	10	10	0,600	0,400
1,5	0,2	10	10	1,200	0,800
1,5	0,3	10	10	1,800	1,200
1,5	0,4	10	10	2,400	1,600
1,6	0,1	10	10	0,615	0,385
1,6	0,2	10	10	1,231	0,769
1,6	0,3	10	10	1,846	1,154
1,6	0,4	10	10	2,462	1,538
1,7	0,1	10	10	0,630	0,370
1,7	0,2	10	10	1,259	0,741
1,7	0,3	10	10	1,889	1,111
1,7	0,4	10	10	2,519	1,481
1,8	0,1	10	10	0,643	0,357
1,8	0,2	10	10	1,286	0,714
1,8	0,3	10	10	1,929	1,071
1,8	0,4	10	10	2,571	1,429
1,9	0,1	10	10	0,655	0,345
1,9	0,2	10	10	1,310	0,690
1,9	0,3	10	10	1,966	1,034
1,9	0,4	10	10	2,621	1,379
2	0,1	10	10	0,667	0,333
2	0,2	10	10	1,333	0,667
2	0,3	10	10	2,000	1,000
2	0,4	10	10	2,667	1,333
2,1	0,1	10	10	0,677	0,323
2,1	0,2	10	10	1,355	0,645
2,1	0,3	10	10	2,032	0,968
2,1	0,4	10	10	2,710	1,290
2,2	0,1	10	10	0,688	0,313

k	λ	n1	n2	σ_2^2	σ_1^2
2,2	0,2	10	10	1,375	0,625
2,2	0,3	10	10	2,063	0,938
2,2	0,4	10	10	2,750	1,250
2,3	0,1	10	10	0,697	0,303
2,3	0,2	10	10	1,394	0,606
2,3	0,3	10	10	2,091	0,909
2,3	0,4	10	10	2,788	1,212
2,4	0,1	10	10	0,706	0,294
2,4	0,2	10	10	1,412	0,588
2,4	0,3	10	10	2,118	0,882
2,4	0,4	10	10	2,824	1,176
2,5	0,1	10	10	0,714	0,286
2,5	0,2	10	10	1,429	0,571
2,5	0,3	10	10	2,143	0,857
2,5	0,4	10	10	2,857	1,143
2,6	0,1	10	10	0,722	0,278
2,6	0,2	10	10	1,444	0,556
2,6	0,3	10	10	2,167	0,833
2,6	0,4	10	10	2,889	1,111
2,7	0,1	10	10	0,730	0,270
2,7	0,2	10	10	1,459	0,541
2,7	0,3	10	10	2,189	0,811
2,7	0,4	10	10	2,919	1,081
2,8	0,1	10	10	0,737	0,263
2,8	0,2	10	10	1,474	0,526
2,8	0,3	10	10	2,211	0,789
2,8	0,4	10	10	2,947	1,053
2,9	0,1	10	10	0,744	0,256
2,9	0,2	10	10	1,487	0,513
2,9	0,3	10	10	2,231	0,769
2,9	0,4	10	10	2,974	1,026
3	0,1	10	10	0,750	0,250
3	0,2	10	10	1,500	0,500
3	0,3	10	10	2,250	0,750
3	0,4	10	10	3,000	1,000
3,1	0,1	10	10	0,756	0,244
3,1	0,2	10	10	1,512	0,488
3,1	0,3	10	10	2,268	0,732
3,1	0,4	10	10	3,024	0,976
3,2	0,1	10	10	0,762	0,238
3,2	0,2	10	10	1,524	0,476
3,2	0,3	10	10	2,286	0,714
3,2	0,4	10	10	3,048	0,952
3,3	0,1	10	10	0,767	0,233
3,3	0,2	10	10	1,535	0,465
3,3	0,3	10	10	2,302	0,698
3,3	0,4	10	10	3,070	0,930
3,4	0,1	10	10	0,773	0,227
3,4	0,2	10	10	1,545	0,455
3,4	0,3	10	10	2,318	0,682
3,4	0,4	10	10	3,091	0,909
3,5	0,1	10	10	0,778	0,222
3,5	0,2	10	10	1,556	0,444

k	λ	n1	n2	σ_2^2	σ_1^2
3,5	0,3	10	10	2,333	0,667
3,5	0,4	10	10	3,111	0,889
3,6	0,1	10	10	0,783	0,217
3,6	0,2	10	10	1,565	0,435
3,6	0,3	10	10	2,348	0,652
3,6	0,4	10	10	3,130	0,870
3,7	0,1	10	10	0,787	0,213
3,7	0,2	10	10	1,574	0,426
3,7	0,3	10	10	2,362	0,638
3,7	0,4	10	10	3,149	0,851
3,8	0,1	10	10	0,792	0,208
3,8	0,2	10	10	1,583	0,417
3,8	0,3	10	10	2,375	0,625
3,8	0,4	10	10	3,167	0,833
4	0,1	10	10	0,800	0,200
1	0,1	15	15	0,750	0,750
1	0,2	15	15	1,500	1,500
1	0,3	15	15	2,250	2,250
1	0,4	15	15	3,000	3,000
1,1	0,1	15	15	0,786	0,714
1,1	0,2	15	15	1,571	1,429
1,1	0,3	15	15	2,357	2,143
1,1	0,4	15	15	3,143	2,857
1,2	0,1	15	15	0,818	0,682
1,2	0,2	15	15	1,636	1,364
1,2	0,3	15	15	2,455	2,045
1,2	0,4	15	15	3,273	2,727
1,3	0,1	15	15	0,848	0,652
1,3	0,2	15	15	1,696	1,304
1,3	0,3	15	15	2,543	1,957
1,3	0,4	15	15	3,391	2,609
1,4	0,1	15	15	0,875	0,625
1,4	0,2	15	15	1,750	1,250
1,4	0,3	15	15	2,625	1,875
1,4	0,4	15	15	3,500	2,500
1,5	0,1	15	15	0,900	0,600
1,5	0,2	15	15	1,800	1,200
1,5	0,3	15	15	2,700	1,800
1,5	0,4	15	15	3,600	2,400
1,6	0,1	15	15	0,923	0,577
1,6	0,2	15	15	1,846	1,154
1,6	0,3	15	15	2,769	1,731
1,6	0,4	15	15	3,692	2,308
1,7	0,1	15	15	0,944	0,556
1,7	0,2	15	15	1,889	1,111
1,7	0,3	15	15	2,833	1,667
1,7	0,4	15	15	3,778	2,222
1,8	0,1	15	15	0,964	0,536
1,8	0,2	15	15	1,929	1,071
1,8	0,3	15	15	2,893	1,607
1,8	0,4	15	15	3,857	2,143
1,9	0,1	15	15	0,983	0,517
1,9	0,2	15	15	1,966	1,034

k	λ	n1	n2	σ_2^2	σ_1^2
1,9	0,3	15	15	2,948	1,552
1,9	0,4	15	15	3,931	2,069
2	0,1	15	15	1,000	0,500
2	0,2	15	15	2,000	1,000
2	0,3	15	15	3,000	1,500
2	0,4	15	15	4,000	2,000
2,1	0,1	15	15	1,016	0,484
2,1	0,2	15	15	2,032	0,968
2,1	0,3	15	15	3,048	1,452
2,1	0,4	15	15	4,065	1,935
2,2	0,1	15	15	1,031	0,469
2,2	0,2	15	15	2,063	0,938
2,2	0,3	15	15	3,094	1,406
2,2	0,4	15	15	4,125	1,875
2,3	0,1	15	15	1,045	0,455
2,3	0,2	15	15	2,091	0,909
2,3	0,3	15	15	3,136	1,364
2,3	0,4	15	15	4,182	1,818
2,4	0,1	15	15	1,059	0,441
2,4	0,2	15	15	2,118	0,882
2,4	0,3	15	15	3,176	1,324
2,4	0,4	15	15	4,235	1,765
2,5	0,1	15	15	1,071	0,429
2,5	0,2	15	15	2,143	0,857
2,5	0,3	15	15	3,214	1,286
2,5	0,4	15	15	4,286	1,714
2,6	0,1	15	15	1,083	0,417
2,6	0,2	15	15	2,167	0,833
2,6	0,3	15	15	3,250	1,250
2,6	0,4	15	15	4,333	1,667
2,7	0,1	15	15	1,095	0,405
2,7	0,2	15	15	2,189	0,811
2,7	0,3	15	15	3,284	1,216
2,7	0,4	15	15	4,378	1,622
2,8	0,1	15	15	1,105	0,395
2,8	0,2	15	15	2,211	0,789
2,8	0,3	15	15	3,316	1,184
2,8	0,4	15	15	4,421	1,579
2,9	0,1	15	15	1,115	0,385
2,9	0,2	15	15	2,231	0,769
2,9	0,3	15	15	3,346	1,154
2,9	0,4	15	15	4,462	1,538
3	0,1	15	15	1,125	0,375
3	0,2	15	15	2,250	0,750
3	0,3	15	15	3,375	1,125
3	0,4	15	15	4,500	1,500
3,1	0,1	15	15	1,134	0,366
3,1	0,2	15	15	2,268	0,732
3,1	0,3	15	15	3,402	1,098
3,1	0,4	15	15	4,537	1,463
3,2	0,1	15	15	1,143	0,357
3,2	0,2	15	15	2,286	0,714
3,2	0,3	15	15	3,429	1,071

k	λ	n1	n2	σ^2_2	σ^2_1
3,2	0,4	15	15	4,571	1,429
3,3	0,1	15	15	1,151	0,349
3,3	0,2	15	15	2,302	0,698
3,3	0,3	15	15	3,453	1,047
3,3	0,4	15	15	4,605	1,395
3,4	0,1	15	15	1,159	0,341
3,4	0,2	15	15	2,318	0,682
3,4	0,3	15	15	3,477	1,023
3,4	0,4	15	15	4,636	1,364
3,5	0,1	15	15	1,167	0,333
3,5	0,2	15	15	2,333	0,667
3,5	0,3	15	15	3,500	1,000
3,5	0,4	15	15	4,667	1,333
3,6	0,1	15	15	1,174	0,326
3,6	0,2	15	15	2,348	0,652
3,6	0,3	15	15	3,522	0,978
3,6	0,4	15	15	4,696	1,304
3,7	0,1	15	15	1,181	0,319
3,7	0,2	15	15	2,362	0,638
3,7	0,3	15	15	3,543	0,957
3,7	0,4	15	15	4,723	1,277
3,8	0,1	15	15	1,188	0,313
3,8	0,2	15	15	2,375	0,625
3,8	0,3	15	15	3,563	0,938
3,8	0,4	15	15	4,750	1,250
4	0,1	15	15	1,200	0,300

ANEXO 2. Resultados parciales para el tamaño de muestra 25

k	λ	P bootstrap	P t Student	P Aleatorizadas
1	0,1	0,1164	0,8339527	0,0744
1,1		0,0574	0,909203	0,113
1,2		0,1546	0,7537723	0,3132
1,3		0,4434	0,6181565	0,1276
1,4		0,669	0,03431782	0,0808
1,5		0,0532	0,7452326	0,0538
1,6		0,2498	0,2142713	0,669
1,7		0,2722	0,2576586	0,0514
1,8		0,6934	0,2721962	0,0624
1,9		0,5354	0,4691766	0,0512
2		0,0674	0,7796078	0,0792
2,1		0,2996	0,6369523	0,0752
2,2		0,0674	0,5442433	0,0672
2,3		0,1318	0,5737821	0,202
2,4		0,0644	0,2205752	0,0684
2,5		0,0834	0,6498982	0,1032

k	λ	P bootstrap	P t Student	P Aleatorizadas
2,6		0,271	0,3144475	0,197
2,7		0,0534	0,7815944	0,364
2,8		0,0782	0,9948408	0,561
2,9		0,1208	0,6782799	0,1858
3		0,0512	0,3183388	0,1756
3,1		0,1578	0,1405343	0,0602
3,2		0,0588	0,2796289	0,0676
3,3		0,6962	0,01766087	0,1208
3,4		0,1078	0,09553761	0,2306
3,5		0,5066	0,2308261	0,3186
3,6		0,085	0,6358834	0,0446
3,7		0,0734	0,2492929	0,0542
3,8		0,0586	0,2013823	0,0794
3,9		0,1088	0,3878859	0,2258
4		0,322	0,4195405	0,3982
1	0,2	0,1284	0,1725083	0,1514
1,1		0,1066	0,2522954	0,066
1,2		0,2072	0,7826592	0,1334
1,3		0,4482	0,7954404	0,0488
1,4		0,1582	0,7566676	0,058
1,5		0,131	0,8142231	0,1116
1,6		0,0674	0,4232385	0,1674
1,7		0,0522	0,5268174	0,2754
1,8		0,0822	0,3198404	0,138
1,9		0,055	0,4373852	0,2014
2		0,1432	0,7151599	0,1882
2,1		0,2918	0,8528832	0,4004
2,2		0,1236	0,9917648	0,2992
2,3		0,1012	0,5160007	0,0678
2,4		0,5892	0,4575493	0,0548
2,5		0,3066	0,4336135	0,0622
2,6		0,0564	0,9993154	0,266
2,7		0,057	0,1829279	0,0612
2,8		0,0544	0,02192216	0,406
2,9		0,053	0,9436274	0,162
3		0,105	0,2929417	0,2032
3,1		0,3596	0,2605077	0,0824
3,2		0,4632	0,1514349	0,21
3,3		0,2264	0,3992088	0,3064
3,4		0,0916	0,941495	0,0594
3,5		0,1192	0,2260919	0,1064

k	λ	P bootstrap	P t Student	P Aleatorizadas
3,6		0,0546	0,649169	0,101
3,7		0,0614	0,6783557	0,0992
3,8		0,0684	0,7240645	0,0576
3,9		0,0872	0,3823939	0,219
4		0,0782	0,4647888	0,0912
1	0,3	0,108	0,8804385	0,1604
1,1		0,1382	0,8708345	0,066
1,2		0,3982	0,8875022	0,218
1,3		0,0774	0,2720648	0,0516
1,4		0,6912	0,8683812	0,0486
1,5		0,1648	0,54936	0,102
1,6		0,078	0,668546	0,1106
1,7		0,2138	0,7489003	0,0578
1,8		0,1492	0,01671687	0,1392
1,9		0,6682	0,01605751	0,308
2		0,1848	0,2920679	0,2988
2,1		0,153	0,4592876	0,0818
2,2		0,1116	0,1196034	0,053
2,3		0,0522	0,2002491	0,2474
2,4		0,1216	0,6820968	0,3484
2,5		0,0782	0,247949	0,0574
2,6		0,2164	0,3597146	0,1882
2,7		0,0526	0,6807238	0,1282
2,8		0,1588	0,03746829	0,0546
2,9		0,0504	0,4959486	0,2118
3		0,3522	0,2790984	0,13
3,1		0,8148	0,1316873	0,1848
3,2		0,2762	0,8259441	0,0538
3,3		0,1138	0,2349045	0,2196
3,4		0,219	0,6869881	0,1228
3,5		0,0774	0,6897043	0,0514
3,6		0,0912	0,7770686	0,2128
3,7		0,1816	0,3174559	0,1364
3,8		0,1036	0,2826701	0,1174
3,9		0,5566	0,4193665	0,0568
4		0,0456	0,6852328	0,0672
1	0,4	0,6156	0,6395685	0,0538
1,1		0,1526	0,6016659	0,0468
1,2		0,1768	0,9519426	0,0818
1,3		0,1198	0,605198	0,0586
1,4		0,0684	0,06113693	0,1126

k	λ	P bootstrap	P t Student	P Aleatorizadas
1,5		0,5044	0,5565297	0,096
1,6		0,0724	0,4922947	0,0564
1,7		0,1638	0,0549049	0,1386
1,8		0,2926	0,1701936	0,3374
1,9		0,1094	0,4596989	0,118
2		0,0498	0,348575	0,215
2,1		0,0608	0,5002974	0,0522
2,2		0,0926	0,8001362	0,083
2,3		0,0794	0,3290155	0,1684
2,4		0,377	0,9184357	0,057
2,5		0,0878	0,684321	0,3094
2,6		0,0622	0,3825343	0,0528
2,7		0,1556	0,262081	0,09
2,8		0,0856	0,4114931	0,0638
2,9		0,1086	0,8366583	0,0576
3		0,055	0,2217663	0,354
3,1		0,3856	0,8146855	0,2012
3,2		0,1456	0,4688127	0,3934
3,3		0,1154	0,3410228	0,065
3,4		0,1024	0,7947751	0,161
3,5		0,0492	0,546698	0,054
3,6		0,0924	0,394429	0,2428
3,7		0,1324	0,4424647	0,1016
3,8		0,2094	0,008294745	0,0608
3,9		0,4992	0,1357176	0,0644
4		0,0652	0,50535	0,0708