



**VICERRECTORADO DE INVESTIGACION, INNOVACION Y  
TRANSFERENCIA DE TECNOLOGIA**

**CENTRO DE POSGRADOS**

**MAESTRÍA EN ENSEÑANZA DE LA MATEMÁTICA**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO  
DE MAGÍSTER EN ENSEÑANZA DE LA MATEMÁTICA**

**TEMA: APLICACIÓN DE UNA NUEVA METODOLOGÍA BASADA EN EL  
MÉTODO BOOTSTRAP PARA LA DETECCIÓN DE RESULTADOS  
ATÍPICOS EN ESTUDIOS INTERLABORATORIO: UN CASO PRÁCTICO  
EN LABORATORIOS CLÍNICOS.**

**AUTOR: BOSSANO CUEVA, JOSÉ ROBERTO.**

**DIRECTOR: M.Sc. FLORES SÁNCHEZ, MIGUEL ALFONSO.**

**SANGOLQUÍ**

**2019**



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA**

**CENTRO DE POSGRADOS**

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, **“APLICACIÓN DE UNA NUEVA METODOLOGÍA BASADA EN EL MÉTODO BOOTSTRAP PARA LA DETECCIÓN DE RESULTADOS ATÍPICOS EN ESTUDIOS INTERLABORATORIO: UN CASO PRÁCTICO EN LABORATORIOS CLÍNICOS”** fue realizado por el señor **BOSSANO CUEVA, JOSÉ ROBERTO** el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que los sustente públicamente.

Sangolquí, 25 de marzo de 2019

Firma:

A handwritten signature in blue ink, which appears to be 'Miguel Flores Sánchez', is written over a faint, light blue circular stamp or watermark.

MSc. Miguel Flores Sánchez

CC: 0918863218



## VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

### CENTRO DE POSGRADOS

#### AUTORÍA DE RESPONSABILIDAD

Yo, **BOSSANO CUEVA, JOSÉ ROBERTO**, con cédula de ciudadanía 1003053293, declaro que el contenido, ideas y criterios del trabajo de titulación: **“APLICACIÓN DE UNA NUEVA METODOLOGÍA BASADA EN EL MÉTODO BOOTSTRAP PARA LA DETECCIÓN DE RESULTADOS ATÍPICOS EN ESTUDIOS INTERLABORATORIO: UN CASO PRÁCTICO EN LABORATORIOS CLÍNICOS”**, es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación es veraz.

Sangolquí, 25 de marzo de 2018

Firma:

A handwritten signature in blue ink, which appears to read 'R. Bossano', is written over a horizontal line.

José Roberto Bossano Cueva

CC: 1003053293



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA**

**CENTRO DE POSGRADOS**

**AUTORIZACIÓN**

Yo, **BOSSANO CUEVA, JOSÉ ROBERTO** con cédula de ciudadanía 1003053293, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación “**APLICACIÓN DE UNA NUEVA METODOLOGÍA BASADA EN EL MÉTODO BOOTSTRAP PARA LA DETECCIÓN DE RESULTADOS ATÍPICOS EN ESTUDIOS INTERLABORATORIO: UN CASO PRÁCTICO EN LABORATORIOS CLÍNICOS**” en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad .

Sangolquí, 25 de marzo de 2018

Firma:

José Roberto Bossano Cueva

CC: 1003053293

## **DEDICATORIA**

*A Gina mi abuelita, a Sofía mi madre, a Washinton mi padre y Carla mi hermana, a quienes amo.*

## **AGRADECIMIENTOS**

A mi director de tesis Miguel Flores por haberme iniciado en este maravilloso tema de manera tan generosa y por haberme apoyado siempre.

Al Ing. Patricio Pugarín por todos los esfuerzos que hizo como coordinador de la maestría para que podamos culminar exitosamente la misma.

A todos los profesores de la Memat III por haber compartido su conocimiento, en especial a la M.Sc. Margarita Kostikova y al Dr. Juan Mayorga.

# ÍNDICE DE CONTENIDOS

<b>CERTIFICACIÓN</b>	<b>i</b>
<b>AUTORÍA DE RESPONSABILIDAD</b>	<b>ii</b>
<b>AUTORIZACIÓN</b>	<b>iii</b>
<b>DEDICATORIA</b>	<b>iv</b>
<b>AGRADECIMIENTOS</b>	<b>v</b>
<b>ÍNDICE DE CONTENIDOS</b>	<b>vi</b>
<b>ÍNDICE DE TABLAS</b>	<b>ix</b>
<b>ÍNDICE DE FIGURAS</b>	<b>x</b>
<b>RESUMEN</b>	<b>xi</b>
<b>ABSTRACT</b>	<b>xii</b>
<b>1 El Problema</b>	<b>1</b>
1.1 Planteamiento del problema . . . . .	1
1.2 Antecedentes del problema . . . . .	2
1.3 Formulación del problema . . . . .	3
1.4 Objetivos . . . . .	4
1.4.1 Objetivo General . . . . .	4
1.4.2 Objetivos específicos . . . . .	5
1.5 Justificación . . . . .	5
<b>2 Los Estudios Interlaboratorio</b>	<b>6</b>
2.1 Introducción . . . . .	6
2.2 Los Estudios para determinar la precisión de un método de medición . .	6

2.3	Organización de un Estudio Interlaboratorio . . . . .	8
2.4	Análisis estadístico de un Estudio de Precisión . . . . .	9
2.5	Estadísticos para el análisis de la consistencia de los resultados . . . . .	10
2.5.1	El Estadístico $h$ de Mandel . . . . .	10
2.5.2	El Estadístico $k$ de Mandel . . . . .	12
2.6	Tratamiento de resultados atípicos . . . . .	13
2.6.1	El Test de Cochran . . . . .	13
2.6.2	El Test de Grubbs . . . . .	15
2.7	Cálculo de la desviación estándar de repetibilidad y reproducibilidad . . . . .	16
<b>3</b>	<b>El Método Bootstrap</b>	<b>20</b>
3.1	Introducción . . . . .	20
3.2	El procedimiento Bootstrap . . . . .	20
3.3	El método Monte Carlo . . . . .	22
3.4	La distribución bootstrap . . . . .	23
3.5	El método bootstrap para determinar diferencias entre grupos . . . . .	24
<b>4</b>	<b>Metodología</b>	<b>26</b>
4.1	Introducción . . . . .	26
4.2	Descripción de la metodología propuesta . . . . .	27
4.3	Hipótesis de reproducibilidad . . . . .	27
4.4	Hipótesis de repetibilidad . . . . .	28
4.5	Algoritmo para la detección de datos atípicos . . . . .	28
4.6	Validación de la metodología propuesta . . . . .	29
4.6.1	La distribución normal . . . . .	31
4.6.2	La distribución de Laplace . . . . .	31
4.6.3	La distribución sesgada . . . . .	31
4.7	Descripción del estudio de simulación . . . . .	31
4.7.1	Incumplimiento de la hipótesis de reproducibilidad . . . . .	32



4.7.2	Incumplimiento de la hipótesis de repetibilidad . . . . .	33
<b>5</b>	<b>Estudio de Simulación</b>	<b>37</b>
5.1	Introducción . . . . .	37
5.2	Proporción de rechazo de la hipótesis nula de reproducibilidad . . . . .	37
5.2.1	Incumplimiento de la hipótesis nula de reproducibilidad para la distribución normal . . . . .	38
5.2.2	Incumplimiento de la hipótesis nula de reproducibilidad para la distribución de Laplace . . . . .	41
5.2.3	Incumplimiento de la hipótesis nula de reproducibilidad para la distribución sesgada . . . . .	44
5.3	Proporción de rechazo de la hipótesis nula de repetibilidad . . . . .	46
5.3.1	Incumplimiento de la hipótesis nula de repetibilidad para la dis- tribución normal . . . . .	47
5.3.2	Incumplimiento de la hipótesis nula de repetibilidad para la dis- tribución de Laplace . . . . .	50
5.3.3	Incumplimiento de la hipótesis nula de repetibilidad para la dis- tribución sesgada . . . . .	53
<b>6</b>	<b>Aplicación de la metodología propuesta</b>	<b>56</b>
6.1	Estudio Interlaboratorio para evaluar la precisión del método empleado para realizar exámenes sanguíneos de Biometría Hemática . . . . .	56
6.1.1	Estudio Interlaboratorio realizado por el método tradicional . . . . .	57
6.1.2	Estudio Interlaboratorio realizado por el método <i>bootstrap</i> . . . . .	65
<b>7</b>	<b>Conclusiones y recomendaciones</b>	<b>72</b>

## ÍNDICE DE TABLAS

Tabla 1	<i>Potencia del estadístico <math>h</math> para la distribución normal</i>	38
Tabla 2	<i>Potencia del estadístico <math>h</math> para la distribución de Laplace</i>	41
Tabla 3	<i>Potencia del estadístico <math>h</math> para la distribución distribución sesgada.</i>	44
Tabla 4	<i>Potencia del estadístico <math>k</math> para la distribución normal</i>	47
Tabla 5	<i>Potencia del estadístico <math>k</math> para la distribución de Laplace</i>	50
Tabla 6	<i>Potencia del estadístico <math>k</math> para la distribución sesgada</i>	53
Tabla 7	<i>Estadísticos <math>h</math> de Mandel</i>	58
Tabla 8	<i>Estadísticos <math>k</math> de Mandel.</i>	61
Tabla 9	<i>Desviaciones estándar de repetibilidad y reproducibilidad</i>	64
Tabla 10	<i>Valores críticos de los estadísticos <math>h</math> bootstrap.</i>	66
Tabla 11	<i>Estadísticos <math>h</math> bootstrap</i>	67
Tabla 12	<i>Valores críticos del estadístico <math>k</math> bootstrap</i>	69
Tabla 13	<i>Estadísticos <math>k</math> bootstrap</i>	70

## ÍNDICE DE FIGURAS

<i>Figura 1</i>	Variaciones en la media y la desviación estándar para la distribución normal.	33
<i>Figura 2</i>	Variaciones en la media y el factor de escala para la distribución de Laplace.	34
<i>Figura 3</i>	Variaciones en la media para la distribución sesgada. . . . .	35
<i>Figura 4</i>	Variaciones en la desviación estándar para la distribución sesgada. . . . .	36
<i>Figura 5</i>	Curvas de potencia del estadístico $h$ para la distribución normal. . . . .	39
<i>Figura 6</i>	La distribución bootstrap y los valores $h$ críticos para la distribución normal.	40
<i>Figura 7</i>	Curvas de potencia del estadístico $h$ para la distribución de Laplace. . . . .	42
<i>Figura 8</i>	La distribución bootstrap y valores $h$ críticos para la distribución de Laplace.	43
<i>Figura 9</i>	Curvas de potencia del estadístico $h$ para la distribución sesgada. . . . .	45
<i>Figura 10</i>	La distribución bootstrap y los valores $h$ críticos para la distribución sesgada.	46
<i>Figura 11</i>	Curvas de potencia del estadístico $k$ para la distribución normal. . . . .	48
<i>Figura 12</i>	La distribución bootstrap y el valor $k$ crítico para la distribución normal. . . . .	49
<i>Figura 13</i>	Curvas de potencia del estadístico $k$ para la distribución de Laplace. . . . .	51
<i>Figura 14</i>	Distribución bootstrap y el valor $k$ crítico para la distribución de Laplace.	52
<i>Figura 15</i>	Curvas de potencia del estadístico $k$ para la distribución sesgada. . . . .	54
<i>Figura 16</i>	La distribución bootstrap y el valor $k$ crítico para la distribución sesgada.	55
<i>Figura 17</i>	Diagrama de barras de los estadísticos $h$ de Mandel. . . . .	59
<i>Figura 18</i>	Diagramas de caja de los estadísticos $h$ de Mandel. . . . .	60
<i>Figura 19</i>	Diagrama de barras de los Estadísticos $k$ de Mandel. . . . .	62
<i>Figura 20</i>	Diagramas de caja de los estadísticos $k$ de Mandel. . . . .	63
<i>Figura 21</i>	Diagramas de caja de los datos provenientes de los laboratorios $E$ , $F$ y $K$ .	65
<i>Figura 22</i>	Valores $h$ críticos bootstrap y los laboratorios 1, y 2 detectados como atípicos. . . . .	68
<i>Figura 23</i>	Valores $k$ críticos bootstrap y los laboratorios 1 y 2 detectados como atípicos. . . . .	71

## RESUMEN

En la presente investigación, se propone el desarrollo de una nueva metodología para detectar datos atípicos en Estudios Interlaboratorio (ILS) basada en el método bootstrap. Para ello, se desarrolló un algoritmo que empleó el método bootstrap no paramétrico para calcular los valores críticos de los estadísticos  $h$  y  $k$  de Mandel, que son utilizados en un contexto ILS para la detección de datos atípicos. Tradicionalmente, durante un Estudios Interlaboratorio se asume que los datos obtenidos por todos los laboratorios siguen la distribución normal. Nosotros desarrollamos un algoritmo bootstrap que no depende del principio de normalidad, en lugar de ello, se calcularon las distribuciones bootstrap de los estadísticos  $h$  y  $k$  de Mandel en base a técnicas de remuestreo aplicadas sobre los datos de una sola muestra. El algoritmo bootstrap propuesto fue evaluado a través de un estudio de simulación, en el cual, se comprobó que los resultados obtenidos por el método propuesto fueron mejores que los obtenidos por el método tradicional cuando los datos provienen de una distribución sesgada y el tamaño de la muestra era pequeño. Finalmente, el algoritmo bootstrap fue aplicado exitosamente a un conjunto de datos que fueron obtenidos técnica computacional de biometría hemática entre laboratorios clínicos.

### **PALABRAS CLAVE:**

- **ESTUDIOS INTERLABORATORIO.**
- **ESTADÍSTICOS  $H$  Y  $K$  DE MANDEL.**
- **MÉTODO BOOTSTRAP.**

## ABSTRACT

In this research, a new methodology for outlier detection in Interlaboratory Studies (ILS) based on the bootstrap methodology is proposed. For this, we developed an algorithm that uses the non-parametric bootstrap to calculate the critical values of Mandel's  $h$  and  $k$  statistics, which are used in an ILS context for outlier detection. Traditionally, an Interlaboratory Study assumes that the data from all the laboratories comes from the normal distribution. We developed a bootstrap algorithm that does not depend of the normality assumption, instead, the bootstrap distributions of the Mandel's  $h$  and  $k$  statistics were calculated based on resampling techniques applied to a single sample data. The proposed bootstrap algorithm was tested through a simulation study, in which, it was found that the results obtained by the proposed method were better than those obtained by the traditional method when data come from a skewed distribution and the sample size was small. Finally, the bootstrap algorithm was successfully applied to a data set that was obtained from a real Interlaboratory Study that had the aim to evaluate computational blood count techniques between clinical laboratories.

### KEY WORDS:

- **INTERLABORATORY STUDIES.**
- **MANDEL'S  $H$  AND  $K$  STATISTICS.**
- **BOOTSTRAP METHOD.**

# 1 El Problema

## 1.1 Planteamiento del problema

Un problema al que se enfrentan las entidades encargadas de estandarizar procesos de medición, como ASTM (*American Society for Testing and Materials*) o ISO (*Organization for Standardization*), es crear normas y protocolos que permitan a las industrias determinar el grado de variabilidad que se pueda presentar en los resultados de las mediciones que se realizan como parte del proceso de control de calidad.

Muchas industrias utilizan laboratorios para realizar ensayos con el fin de verificar el cumplimiento de ciertos requisitos de calidad sobre los materiales que se emplean, ante lo cual, es necesario contar con procesos de medición que arrojen resultados exactos y confiables. El área encargada de estudiar la variación presente en los procesos de medición dentro de los laboratorios de ensayo, son los estudios de Repetibilidad y Reproducibilidad (r&R).

La repetibilidad se define como “la variación que se presenta en una medición causada por el instrumento de medida empleado, ya que las pruebas las obtiene un único operario en cada laboratorio con el mismo instrumento y sin que transcurra demasiado tiempo entre varias mediciones” (Pastoriza, 2013). Por otra parte, la reproducibilidad es “la variación debida a los laboratorios encargados de aplicar un método de medición” (Barrentine, 2003). Es decir, la reproducibilidad ayuda a establecer el error o “con esto se puede determinar el sesgo que se produce debido a los operarios o trabajadores del laboratorio en cuestión” (Pastoriza, 2013).

La variabilidad presente en un proceso de medición, asumiendo que dicho proceso se halla bajo control, puede deberse a factores naturales como cambios en la temperatura del ambiente, deficiente apreciación visual de los operarios, entre otros. En caso de la variabilidad sea producida por causas diferentes a las mencionadas: uso de equipos descalibrados, o incorrecta aplicación de un método de medición, se puede emplear los estudios de repetibilidad y reproducibilidad para determinar la

variabilidad del proceso de medición. En este caso, si las mediciones hechas por un laboratorio son muy diferentes comparadas con otras, el laboratorio en cuestión deberá ser investigado.

Con el objetivo de detectar las causas de la variabilidad presente en al aplicar un método de medición, y como parte de los estudios r&R, surgieron los Estudios Interlaboratorio, los mismos que evalúan “las consistencias de los resultados obtenidos cuando un grupo de laboratorios aplican un método bien definido de medición, realizando mediciones sucesivas sobre el mismo material de prueba” (Highways and Transportation, 1996).

## **1.2 Antecedentes del problema**

El procedimiento para realizar Estudios Interlaboratorio se encuentra detallado en las normas ISO 5725 (*Accuracy - Trueness and precision - of measurement methods and results*), y ASTM E-691 (*Standard practice for conducting an interlaboratory study to determine the precision of a test method*).

El Análisis de resultados de un Estudio Interlaboratorio se lleva a cabo mediante una prueba de Análisis de Varianza de Efectos Fijos (ANOVA), la cual requiere que los datos sigan una distribución normal y las varianzas entre laboratorios sean similares. “El procedimiento detallado en las normas ASTM E-691 e ISO 5725-2 no emplea formalmente un ANOVA, sin embargo, llega a los mismos resultados” (Vander Heyden and Smeyers-Verbeke, 2007).

Los valores atípicos o extremos, de forma general, se definen como valores que pertenecen a una distribución diferente a la que pertenecen la mayor parte del conjunto de datos. En el caso de los Estudios Interlaboratorio es sumamente útil la detección de resultados atípicos ya que ello podría ayudar a detectar problemas de índole técnico dentro de un laboratorio puesto que “la presencia de resultados atípicos dentro de un laboratorio, podría ser un síntoma de problemas en la aplicación de un método de medición, lo que debe ser investigado y reportado” (Highways and Transportation,

1996).

Las herramientas propuestas en las normas ISO 5725 y ASTM E-691 para detectar valores atípicos y evaluar la consistencia de los datos, son los estadísticos  $h$  y  $k$  de Mandel. El estadístico  $h$  mide el nivel de consistencia a nivel interlaboratorio, ya que “compara el promedio de los resultados de un laboratorio con el promedio del resto de laboratorios, por lo tanto, se refiere a la discrepancia del valor de la media” (Pastoriza, 2013). Mientras que el estadístico  $k$  mide la consistencia de los resultados a nivel intralaboratorio, “comparando la variabilidad dentro de las mediciones internas de un laboratorio con el resto de laboratorios” (Pastoriza, 2013).

Para poder detectar si una medición es atípica, es necesario calcular los valores críticos de los estadísticos  $h$  (véase en (1)) y  $k$  (véase en (3)) por cada laboratorio. Si el valor absoluto de una puntuación  $h$  o  $k$  es mayor que sus respectivos valores críticos, la medición se considera atípica, y el laboratorio correspondiente es detectado como atípico.

### **1.3 Formulación del problema**

La metodología para detectar valores atípicos incluida en las normas ISO 5725 y ASTM E-691 propone hallar los valores críticos de los estadísticos  $h$  y  $k$  en base a una serie de supuestos teóricos sobre las variables aleatorias que se están estudiando, estos son, básicamente, que los datos sigan la distribución normal.

En la actualidad, el uso de métodos basados en remuestreo ha ido aumentando hasta el punto en que dichos métodos han sido aplicados en diversas áreas, una de las características de estos métodos es que “permiten obtener la distribución muestral de un estadístico, en casos donde la única información que disponemos es una muestra” (Hesterberg, 2003). Precisamente, estos métodos realizan, de manera sucesiva, muestreo de los elementos de una muestra para generar miles de nuevas muestras, para proceder a calcular el valor del estadístico en cada muestra (llamadas remuestras).



La finalidad de estos métodos es obtener una distribución muestral empírica del estadístico.

Los métodos de remuestreo son especialmente útiles en caso donde la única información que disponemos es una muestra y desconocemos la población de la cual proviene. Uno de los tipos de métodos de remuestreo más importantes en la actualidad es el bootstrap debido a sus aplicaciones en diversos campos. En el caso de los Estudios Interlaboratorio este método ha sido aplicado por autores como (Flores, 2018).

El método bootstrap básicamente se encarga de estimar la distribución muestral de un estadístico a partir de una muestra. Su metodología se basa en generar de múltiples muestras con reemplazamiento de los datos de la muestra que se ha recogido en el experimento. La distribución que se genera mediante este método de conoce como distribución bootstrap del estadístico.

La distribución bootstrap es sumamente útil ya que “brinda información sobre el centro, la forma, y la desviación de la distribución de un estadístico” (Hesterberg, 2003), por lo cual, podría convertirse en una herramienta muy útil para determinar la distribución empírica de los estadísticos  $h$  y  $k$ , lo que podría llevar a detectar mediciones atípicas mediante una metodología diferente a la tradicional.

La tarea fundamental de esta investigación será, por lo tanto, desarrollar un algoritmo que utilice métodos de remuestreo, en especial, la metodología bootstrap para detectar resultados atípicos en Estudios Interlaboratorio.

## **1.4 Objetivos**

### **1.4.1 Objetivo General**

Diseñar una nueva metodología basada en el método bootstrap para detectar resultados atípicos en Estudio Interlaboratorios.

### 1.4.2 Objetivos específicos

1. Elaborar en el software R un algoritmo que permita detectar resultados atípicos a través de Métodos de Remuestreo.
2. Diseñar un estudio de simulación para determinar la potencia del nuevo algoritmo para detectar datos atípicos.
3. Realizar un Estudio Interlaboratorio para determinar la precisión del Método Computarizado para realizar Biometrías Hemáticas en Laboratorios Clínicos.
4. Evaluar los resultados de la aplicación de datos reales al algoritmo propuesto.

## 1.5 Justificación

Los valores críticos de los estadísticos  $h$  y  $k$  propuestos en las normas ASTM E 691 e ISO 5725-2 se hallan en base a una serie de supuestos teóricos que se aplican sobre las variables en estudio, lo que implica que se desconozca la verdadera distribución de los estadísticos  $h$  y  $k$  de Mandel. Nuestra propuesta es determinar una aproximación a la distribución real de estos estadísticos mediante la metodología bootstrap, para luego obtener los cuantiles de la distribución bootstrap de dichos estadísticos y con ello poder detectar valores atípicos.

Esta investigación pretende aportar con una nueva metodología para la detección de resultados atípicos que sirva para establecer los alcances del método bootstrap a los Estudios Interlaboratorio, y que beneficie a los laboratorios de industrias nacionales que busquen desarrollar Estudios Interlaboratorio para detectar resultados atípicos de forma más rápida.

## **2 Los Estudios Interlaboratorio**

### **2.1 Introducción**

Los Estudios Interlaboratorio forman parte de los estudios de Repetibilidad y Reproducibilidad, y su principal función es reunir a un grupo de laboratorios para proceder a analizar la misma muestra con el objetivo de comparar los resultados del conjunto de laboratorios y detectar posibles errores en la aplicación de un método de medición, o también, reconocer problemas con la instrumentación de medida empleada por los laboratorios.

Los Estudios Interlaboratorio también pueden emplearse como “investigaciones colaborativas en las cuales un grupo de laboratorios analizan los mismos materiales con el fin de obtener información sobre las características de dichos materiales” (Vander Heyden and Smeyers-Verbeke, 2007).

Dependiendo de los objetivos que se quiera alcanzar con la información recopilada, existen tres tipos de Estudios Interlaboratorio:

- Estudio de la precisión de un método de medición.
- Estudios de proficiencia para acreditar la calidad de laboratorios.
- Estudios para la certificación de materiales.

En esta investigación nos centraremos en los estudios de la precisión de un método de medición.

### **2.2 Los Estudios para determinar la precisión de un método de medición**

Los Estudios Interlaboratorio se utilizan para evaluar la consistencia de las mediciones obtenidas por un grupo de laboratorios “cuando se analiza el mismo material de prueba, y todos los laboratorios participantes aplican el mismo método de medición” (Highways and Transportation, 1996).

Se define a la precisión de un método de medición como “el grado en que tienden a coincidir las mediciones repetidas de la misma cantidad. La precisión se define mediante la desviación estándar del proceso de medición” (Navidi, 2006). En un Estudio Interlaboratorio, “la precisión de un método de medición se determina mediante el cálculo de la repetibilidad y la reproducibilidad” (ISO, 1994). La repetibilidad y reproducibilidad también nos indican las condiciones que deben ser tomadas en cuenta para a la realización de mediciones en un Estudio Interlaboratorio.

Se define a la repetibilidad como la “precisión bajo condiciones de repetibilidad” (ASTM, 2005). Esto implica “condiciones donde mediciones independientes son obtenidas con el mismo método, en idénticos materiales, en el mismo laboratorio, por el mismo operador, usando el mismo instrumento de medida en cortos intervalos de tiempo entre medición y medición” (ASTM, 2005). Es decir, “las mediciones deben realizarse sin recalibración de equipos y aplicando todos los pasos del método de medición en cada ocasión”. (Vander Heyden and Smeyers-Verbeke, 2007).

Por otra parte, se define a la reproducibilidad como la “precisión bajo condiciones de reproducibilidad” (ASTM, 2005), es decir “condiciones donde las mediciones se obtienen con el mismo método aplicado a los mismos materiales, en distintos laboratorios, con diferentes operadores, usando diferentes instrumentos de medida” (ASTM, 2005). Podemos ver que para determinar la reproducibilidad es necesario que exista un trabajo colaborativo entre los laboratorios participantes.

El nivel de significancia se define como “la probabilidad de decidir erróneamente que dos conjuntos de datos provienen de distintas poblaciones” (Highways and Transportation, 1996). El nivel de significancia que sugiere la norma ASTM E-691 es de 0.005, mientras que la norma ISO 5725 sugiere el valor de significancia de 0.010, lo que indica que hay un 0.5%, o 1%, de probabilidad de establecer que dos conjuntos de datos vienen de distintas poblaciones, cuando en realidad provienen de la misma población.

### 2.3 Organización de un Estudio Interlaboratorio

La primera parte de un Estudio Interlaboratorio consiste en la planificación y el diseño de los procedimientos requeridos para llevar a cabo un Estudio Interlaboratorio.

El estudio debe ser organizado y supervisado por “un panel de expertos en la aplicación del método de medición que va a ser evaluado” (Vander Heyden and Smeyers-Verbeke, 2007). Estas personas deben elegir a un coordinador con amplia experiencia en el método de medición que se investigará. De igual manera se debe elegir una persona con conocimientos en estadística que esté en capacidad de analizar e interpretar los resultados respectivos.

Los laboratorios seleccionados “deben ser representativos de la población de laboratorios que aplican un método estándar de medición” (Highways and Transportation, 1996). No es recomendable elegir únicamente laboratorios de alto prestigio y nivel de capacidad, “con el fin de no obtener resultados sesgados” Vander Heyden and Smeyers-Verbeke (2007).

En los Estudios Interlaboratorio es común que “a los laboratorios participantes se les envíen las mismas muestras (homogeneizadas o que los laboratorios deberán homogeneizar) para su análisis, para posteriormente estudiar los resultados conjuntos de dichos análisis y valorar la consistencia de los resultados” (Pastoriza, 2013).

Puesto que en un Estudio Interlaboratorio el objetivo es obtener una estimación  $s$  de la desviación estándar  $\theta$  de la población de laboratorios que emplean el mismo método de medición, se tiene que tomar en cuenta que a mayor cantidad de mediciones, más próximo estará el  $s$  al valor real de  $\theta$ , por lo tanto, se recomienda que el número de laboratorios participantes oscile entre 8 y 15, “aunque otras normas estipulan que se requiere de al menos 15 laboratorios” (Vander Heyden and Smeyers-Verbeke, 2007).

En un Estudio Interlaboratorio, un material, o nivel de medición, se define como “la caracterís-

tica de un cuerpo o sustancia que es factible de ser medida” (Vander Heyden and Smeyers-Verbeke, 2007). En un Estudio Interlaboratorio, se recomienda evaluar al menos al menos cinco materiales dentro de la muestra.

En cuanto al número de mediciones, o réplicas, que se deben realizar para cada material, la norma ASTM E-691 recomienda tomar tres réplicas de cada medición, la norma ISO 5725 no hace referencia a un número de específico de réplicas por laboratorio, pero al menos cada laboratorio debe entregar dos réplicas.

## **2.4 Análisis estadístico de un Estudio de Precisión**

El análisis estadístico de un estudio colaborativo de precisión de un método analítico consta de las siguientes etapas sucesivas según la norma ISO 5725:

1. Exámen crítico de los datos con el objetivo de identificar datos inconsistentes, los cuales, pasarían a convertirse en posibles datos atípicos.
2. Confirmar la existencia de valores atípicos y otras irregularidades.
3. Determinar los valores preliminares de precisión y las medias para cada material separadamente.
4. Establecer los valores finales de precisión y las medias para cada material, incluyendo, de ser posible, el establecimiento de una relación matemática entre los valores de precisión y la media del material.

El modelo estadístico básico para analizar los resultados de un Estudio Interlaboratorio es un Análisis de Efectos Aleatorios de Una Vía (ANOVA), y pese a que las normas ISO y ASTM no apliquen los cálculos requeridos por el modelo ANOVA, llegan a los mismos resultados. Para la aplicación de este modelo, se deben cumplir algunas condiciones: los datos de todos los grupos

deben tener provenir de la distribución normal, y las varianzas entre los grupos deben ser similares.

Una vez que ha concluido el Estudio Interlaboratorio, todos los resultados de las mediciones son enviados a los coordinadores del estudio, quienes entregan los datos al encargado de realizar los cálculos estadísticos para empezar por el proceso de análisis de datos.

## **2.5 Estadísticos para el análisis de la consistencia de los resultados**

El primer paso para detectar datos atípicos en un estudio es calcular los estadísticos  $h$  y  $k$  de Mandel puesto que “son técnicas numéricas y gráficas para evaluar la calidad de los laboratorios” (Vander Heyden and Smeyers-Verbeke, 2007).

Ya que la “eliminación automática de datos atípicos no se aplica en Estudios Interlaboratorio” Vander Heyden and Smeyers-Verbeke (2007), es necesario contar con test gráficos que permitan detectar datos atípicos, para luego proceder a analizar su influencia en el conjunto de datos.

### **2.5.1 El Estadístico $h$ de Mandel**

El estadístico  $h$  expresa la consistencia de los resultados cuando estos se comparan a nivel interlaboratorio. Este estadístico “mide el promedio de los resultados de un laboratorio comparado con el promedio del resto de laboratorio, se refiere entonces a la discrepancia del valor de la media” (Pastoriza, 2013). De esta forma, “el estadístico  $h$  representa una aproximación al sesgo de medición de un laboratorio” (Vander Heyden and Smeyers-Verbeke, 2007), en donde se asume a la media general obtenida por todos los laboratorios, como el valor de referencia de un material.

A continuación se obtiene la expresión para calcular el estadístico  $h$  de Mandel, la información se basa en el reporte de (Wilrich, 2013):

Sea  $(x_1, x_2, \dots, x_L)$  una muestra de  $L$  observaciones, donde  $L$  es el número de laboratorios. Los  $x_l; l = 1, \dots, L$  son modelados como realizaciones de variables aleatorias  $X_l; l = 1, \dots, L$  independientemente e idénticamente distribuidas acorde a una distribución normal  $N(\mu, \sigma^2)$ .

El valor crítico del estadístico  $h$  de Mandel (Wilrich, 2013), se define mediante:

$$h_{l;1-\frac{\alpha}{2}} = \frac{(L-1)t_{L-2;1-\frac{\alpha}{2}}}{\sqrt{L(t_{L-2;1-\frac{\alpha}{2}}^2 + L-2)}} \quad (1)$$

Donde  $t_{L-2;1-\frac{\alpha}{2}}$  es el cuantil  $(1 - \frac{\alpha}{2})$  de la distribución  $t$  con  $v = L - 2$  grados de libertad.

Para el caso cuando  $L$  laboratorios realizan  $n$  réplicas sobre un material, se define el estadístico  $h$  de Mandel de la siguiente forma:

$$h_l = \frac{d_l}{S_{\bar{x}}} = \frac{\bar{x}_l - m}{\sqrt{\sum_{l=1}^L \frac{(\bar{x}_l - m)^2}{L-1}}}; l = 1, \dots, L \quad (2)$$

Donde  $\bar{x}_l$  es la media obtenida entre las réplicas hechas sobre el mismo material (el valor  $\bar{x}_l$  se conoce como media de la celda),  $m$  es el promedio de las medias de las celdas obtenidas por todos los laboratorios (la media de las medias), es decir:

$$\bar{x}_l = \frac{\sum_{i=1}^n x_i}{n}$$

y,

$$m = \frac{\sum_{l=1}^L \bar{x}_l}{L}$$

Se detecta un laboratorio como atípico cuando el valor del estadístico  $h_l$  es mayor que el valor crítico, es decir  $h_l \geq h_{l;1-\frac{\alpha}{2}}$ .

En el gráfico de barras del estadístico  $h$ , se debe indicar los valores críticos positivos y negativos del estadístico mediante líneas horizontales que partan de dichos valores críticos. La norma ISO 5725-2 sugiere que se deben mostrar los valores críticos para los niveles de 0.5 % y de 0.1%. Por



otra parte, la norma ASTM E-691 indica que se debe indicar los valores críticos para el nivel de 0.5 %, y no al nivel de 1% ya que “con el 1% muchas de las mediciones quedan fuera del nivel crítico” (Pastoriza, 2013) y muchos laboratorios son detectados como atípicos.

### 2.5.2 El Estadístico $k$ de Mandel

El Estadístico  $k$  de Mandel “mide la consistencia intralaboratorio, comparando la variabilidad interna de un laboratorio individual con el resto de laboratorios” (Pastoriza, 2013).

A continuación se obtiene la expresión para calcular el estadístico  $k$  de Mandel, la información se basa en el reporte de (Wilrich, 2013):

Sean  $(S_1^2, S_2^2, \dots, S_L^2)$  las  $L$  varianzas muestrales, cada una basada en  $n$  valores observados. Bajo la hipótesis de que los valores de las observaciones  $x_{ji}; j = 1, 2, \dots, L, i = 1, 2, \dots, n$  son realizaciones de variables aleatorias  $X_{ji}$ , las cuales cada  $j$  son idénticamente distribuidas e independientes acorde a una distribución normal  $N(\mu_i, \sigma^2)$ , las varianzas muestrales  $S_j^2; j = 1, \dots, L$  divididas por  $\sigma^2$  siguen una distribución  $\chi^2/v$  con  $v = n - 1$  grados de libertad.

El valor crítico del Estadístico  $k$  de Mandel (Wilrich, 2013), viene dado por

$$k_{l,n;1-\alpha} = \sqrt{\frac{L}{\left(1 + \frac{L-1}{F_{v_1, v_2; \alpha}}\right)}} \quad (3)$$

Donde  $F_{v_1, v_2; \alpha}$  es el  $\alpha$ -cuantil de la distribución  $F$  con  $v_1 = (L - 1)(n - 1)$  y  $v_2 = n - 1$  grados de libertad.

Para el caso cuando  $L$  laboratorios tomen  $n$  réplicas de un material cualquiera, se define el estadístico  $k$  mediante:

$$k_l = \frac{S_i}{S_r} = \frac{\sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}}}{\sqrt{\frac{\sum_{l=1}^L S_l^2}{L}}} \quad (4)$$

Donde  $S_i$  es la desviación estándar del número de repeticiones hechas por el laboratorio en estudio (también llamada desviación estándar intracelda).  $S_r$  se denomina desviación estándar de repetibilidad y es la desviación estándar combinada de todos los laboratorios para ese material.

Se detecta un laboratorio como inconsistente cuando el valor del estadístico  $k_l$  es mayor que el valor crítico, es decir  $k_l \geq k_{l,n;1-\alpha}$ .

Los valores obtenidos del estadístico  $k$  se deben representar en un diagrama de barras, en el cual, se debe indicar el valor crítico del estadístico para un nivel de significancia de 1%, o del 0.5%.

## 2.6 Tratamiento de resultados atípicos

Las normas ASTM E-691 e ISO 5725, definen tres tipos de datos atípicos: mediciones atípicas comparadas con las del mismo laboratorio (tipo 1), medias atípicas comparadas con las del resto de laboratorios (tipo 2), y, desviaciones estándar atípicas comparadas con las del resto de laboratorios (tipo 3). Para los atípicos de tipo 1 y 2, “los mismos test son aplicables” (Vander Heyden and Smeyers-Verbeke, 2007).

La norma ISO 5725 sugiere que deben examinarse los tres tipos de datos atípicos en un Estudio Interlaboratorio. Este análisis se lo realiza mediante el Test de Cochran, para datos atípicos tipo 3, y el Test de Grubbs, para datos atípicos tipo 1 y 2.

### 2.6.1 El Test de Cochran

La norma ISO 5725 indica que “se debe empezar por aplicar de forma sucesiva el Test Cochran de cola superior para la comparación de las varianzas de los laboratorios” (ISO, 1994), luego se

procede a analizar los resultados y “si dos o tres laboratorios muestran desviaciones estándar muy altas, especialmente para un solo material, las conclusiones del Test de Cochran deben ser tomadas con cuidado” (Vander Heyden and Smeyers-Verbeke, 2007). Es decir, no podemos concluir que un laboratorio sea atípico por poseer una desviación estándar alta para un solo material.

Si en nuestro análisis encontramos valores inconsistentes para diferentes materiales en un mismo laboratorio, es una señal de que la varianza intralaboratorio es muy alta, y “todas las mediciones correspondientes a ese laboratorio deben ser omitidas del análisis” (Vander Heyden and Smeyers-Verbeke, 2007).

Para determinar mediante la prueba de Cochran (Cochran, 1941) si la mayor varianza  $S_{max}^2 = \max(S_1^2, S_2^2, \dots, S_L^2)$  es un valor atípico, se utiliza el estadístico de prueba:

$$C = \frac{S_{max}^2}{\sum_{j=1}^L S_j^2} = \frac{1}{L} \max_{i=1, \dots, L} \frac{S_i^2}{\bar{S}^2} \quad (5)$$

Para esta prueba el valor crítico esta dado en (Wilrich, 2013) por la siguiente expresión:

$$c_{l,n;1-\alpha} \approx \frac{1}{(1 + (L - 1)F_{v_1, v_2; \frac{\alpha}{L}})} \quad (6)$$

Donde  $F_{v_1, v_2; \frac{\alpha}{L}}$  es el  $\frac{\alpha}{L}$ -cuantil de la distribución  $F$  con  $v_1 = (L - 1)(n - 1)$  y  $v_2 = n - 1$  grados de libertad.

La prueba de Cochran, es una prueba de valores atípicos de una cola porque examina solamente el valor más alto en una serie de varianzas. Como se dijo anteriormente, si un laboratorio presenta una gran cantidad de valores atípicos para diferentes materiales, el laboratorio es considerado como atípico, y los datos de dicho laboratorio son removidos del estudio.

### 2.6.2 El Test de Grubbs

Luego de la aplicación del Test de Cochran, se debe aplicar el Test de Grubbs para detectar medias atípicas entre los laboratorios.

Si se desea determinar si la observación  $X_{\max} = \max(X_1, \dots, X_n)$  es un valor atípico, se utiliza la prueba de Grubbs (Grubbs, 1950). El estadístico para esta prueba es:

$$G = \max_{i=1, \dots, n} H_i = \max_{i=1, \dots, n} \frac{X_i - \bar{X}}{S} = \frac{X_{\max} - \bar{X}}{S} \quad (7)$$

En el caso que se quiera determinar si la menor observación  $X_{\min} = \min(X_1, \dots, X_n)$  es un valor atípico se usa el estadístico de prueba:

$$G = \max_{i=1, \dots, n} -H_i = \max_{i=1, \dots, n} \frac{X_i - \bar{X}}{S} = \frac{X_{\min} - \bar{X}}{S} \quad (8)$$

El valor crítico (Wilrich, 2013) correspondiente para esta prueba es:

$$g_{n;1-\alpha} \approx \frac{(n-1)t_{n-2;1-\frac{\alpha}{n}}}{\sqrt{n(n-2+t_{n-2;1-\frac{\alpha}{n}}^2)}} \quad (9)$$

Donde  $t_{n-2;1-\frac{\alpha}{n}}$  es el cuantil  $(1 - \frac{\alpha}{n})$  de la distribución  $t$  con  $v = n - 2$  grados de libertad.

Para el caso cuando se tienen  $L$  laboratorios y  $n$  réplicas en cada uno, se define el estadístico  $g_{L;1-\alpha}$  sustituyendo las observaciones por las medias de los resultados de cada laboratorio y la media de las observaciones por la media global de los promedios de los laboratorios.

Si un laboratorio es detectado con resultados atípicos, aplicando el estadístico  $h$ , o la prueba de Grubbs a diferentes niveles dentro de un laboratorio, indicaría que el sesgo del laboratorio es alto. Lo que puede deberse a “un error sistemático, una mala calibración, o errores de ecuación al computar los resultados” (ISO, 1994).

## 2.7 Cálculo de la desviación estándar de repetibilidad y reproducibilidad

El propósito fundamental de los Estudios Interlaboratorio es determinar la precisión de un método de medición mediante el cálculo de las desviaciones de repetibilidad y la reproducibilidad de un método de medición representadas por  $S_r$  y  $S_R$  respectivamente.

La desviación estándar de repetibilidad  $S_r$  se define como la “variación en condiciones de repetibilidad” (ISO, 1994), y debe ir acompañada de otra cantidad conocida como el Límite de repetibilidad  $r$ , que se define como “el valor que se espera no sea excedido por la diferencia de dos mediciones sucesivas de un material, con una probabilidad del 95%” (Highways and Transportation, 1996).

El límite de repetibilidad se define de la siguiente forma:

$$r = 2.8S_r$$

Por otra parte, la desviación de reproducibilidad  $S_R$  se define como la “variación en condiciones de reproducibilidad” (ISO, 1994), y debe ir acompañada del cálculo del límite de Reproducibilidad  $R$ , el cual se define como “el valor que se espera no sea excedido por la diferencia de dos mediciones realizadas bajo condiciones de reproducibilidad, con una probabilidad del 95%” (Highways and Transportation, 1996)

El límite de reproducibilidad se define mediante:

$$r = 2.8S_R$$

Los límites de repetibilidad y reproducibilidad son utilizados para emitir conclusiones sobre la precisión de un método, convirtiéndose en información imprescindible en un informe de un estudio interlaboratorio.

La norma ISO 5725-2, proporciona un modelo estadístico básico por el cual la estimación de la veracidad y la precisión de un método analítico bien dado por:

$$y = m + B + e$$

Donde:

- $y$  es el resultado de la medición.
- $m$  es la media general para el material analizado.
- $B$  es la componente del sesgo del laboratorio.
- $e$  es el error aleatorio que ocurre en cada medición.

El error aleatorio  $e$  es calculado por medio de  $S_r^2$ , la varianza de repetibilidad. Mientras la varianza  $B$  se estima por la varianza interlaboratorio  $S_L^2$ .

La varianza de reproducibilidad  $S_R^2$  se calcula por medio de:

$$S_R^2 = S_r^2 + S_L^2 \quad (10)$$

Para poder calcular dichas desviaciones, es preciso empezar por calcular la media de cada laboratorio en cada nivel.

Al calcular la media (por ejemplo, de las tres mediciones del primer laboratorio para el nivel A), se tendrá que esta es igual a:

$$\hat{y}_i = \frac{1}{n} \sum_{j=1}^{n_i} y_{ij}$$

Donde  $n$  es el número de réplicas hechas de cada medición.

El siguiente paso consiste en expresar la media general para cada nivel de concentración, es decir, el promedio de las medias de cada uno de los  $p$  laboratorios para un determinado nivel.

$$\hat{m} = \frac{\sum_{i=1}^p \hat{y}_i}{p}$$

Una vez que conocemos el valor de la media general de cada nivel, es necesario conocer la dispersión respecto a esa media. Para ello, la variabilidad interna  $S_i$  de cada laboratorio en cada uno de los niveles estudiados, se calcula de la forma que sigue:

$$S_i = \sqrt{\frac{1}{n-1} \sum_{n=1}^n (y_{ij} - \bar{y}_i)^2}$$

Para obtener  $S_r$ , se obtiene la sumatoria de las desviaciones típicas  $S_i$  calculadas anteriormente de cada uno de los niveles en cada laboratorio al cuadrado, se divide para el número de laboratorios y se obtiene la raíz cuadrada. La variabilidad interna promedio  $S_r$  es conocida como el ruido presente en el experimento.

$$S_r = \sqrt{\frac{\sum_{i=1}^n S_i^2}{p}}$$

La desviación típica entre-laboratorios  $S_L$ , se obtiene a partir de la varianza  $S_m^2$  y la varianza de repetibilidad  $S_r^2$ , mediante la ecuación:

$$S_L = \sqrt{S_m^2 + S_r^2 \frac{n-1}{n}}$$

Donde  $S_m^2$  es la varianza de la diferencia entre las medias de las  $n$  lecturas hechas por un laboratorio  $y_i$ , con la media general de cada material  $m$ , es decir:

$$S_m = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (\bar{y}_i - \hat{m})^2}$$

Una vez que  $S_r$  y  $S_L$  han sido calculados, pueden reemplazarse en la expresión (10) y calcular el valor de  $S_R$ , es decir, la varianza de reproducibilidad. Estos valores serán necesarios para calcular los límites de repetibilidad y reproducibilidad como veremos más adelante.



## 3 El Método Bootstrap

### 3.1 Introducción

Este método se enmarca dentro de los métodos de remuestreo, los cuales reciben su nombre debido a que asumen a una sola muestra como la población y a partir de ella generan múltiples nuevas muestras (remuestras) para obtener la función de distribución empírica de la población. Dependiendo de la forma en que se generen las nuevas muestras existen diferentes tipos de métodos de remuestreo.

El método bootstrap es un tipo de método de remuestreo fue desarrollado por Bradley Efron en 1979, y posteriormente este método recibió grandes aportes de otros investigadores siendo el más relevante Robert Tibshirani.

El método bootstrap “es un procedimiento empleado para generar una gran cantidad de nuevas muestras a partir de una sola muestra, con el fin de construir una distribución que se asemeje a la distribución muestral del estadístico” (Hesterberg, 2003). La distribución que se obtiene mediante este método se conoce como Distribución Bootstrap del Estadístico.

El método bootstrap es útil en casos donde la única información que disponemos es una muestra y, además, “en casos donde el tamaño de la muestra es pequeño y no se puede asegurar el cumplimiento del Teorema Central del Límite o el tamaño de la muestra es muy pequeño” (Hesterberg, 2003).

### 3.2 El procedimiento Bootstrap

De manera general, el método bootstrap consta de los siguientes pasos:

1. Se toma una muestra aleatoria  $x_1, \dots, x_n$  y se generan grandes cantidades de muestras a

partir de la muestra original llamadas Muestras Bootstrap o Remuestras, este proceso se lo realiza mediante muestreo con reemplazamiento de los elementos de la muestra original, procurando que cada remuestra sea del mismo tamaño que la muestra original.

2. Se calcula el valor del estadístico en cada muestra Bootstrap. A la distribución generada por el conjunto de estadísticos calculados en cada remuestra se le conoce como Distribución Bootstrap del Estadístico.
3. Se emplea la distribución bootstrap que ha sido estimada para obtener información sobre la forma, la localización, y la dispersión de la distribución muestral del estadístico, puesto que “la distribución Bootstrap de un estadístico, basada en una gran cantidad de remuestras, se aproxima a la distribución muestral del estadístico” (Hesterberg, 2003).

El proceso que describimos es conocido como bootstrap no paramétrico. Sin embargo, existen otras variantes de este método bootstrap, como el bootstrap paramétrico, el cuál asume que la distribución de los datos tiene una forma paramétrica, de manera general, se asume que la función de distribución es “la distribución normal, con el fin de llegar a estimaciones bootstrap, que son los estimadores de máxima verosimilitud” (Figueireido and Gomes, 2013). Existen otros tipos de bootstrap, como el bootstrap suavizado, sin embargo, en la presente investigación nos limitaremos a estudiar el método bootstrap paramétrico y no paramétrico.

El bootstrap paramétrico hace uso del Método de Monte Carlo, ya que “los métodos de Monte Carlo implican el muestreo a partir de una distribución de probabilidad completamente especificada” (Figueireido and Gomes, 2013), al contrario del bootstrap no paramétrico, en el cuál no se especifica la función de distribución.

A continuación mostramos las características del método de Monte Carlo, el cual es empleado

en el bootstrap no paramétrico.

### 3.3 El método Monte Carlo

Aunque formalmente el método Monte Carlo no es un método de remuestreo, si es un método adecuado para la generación de números aleatorios. Ya que “las muestras aleatorias se generan a partir de la simulación estocástica de realizaciones de variables aleatorias que poseen una función de distribución de probabilidad que ha sido especificada previamente” (Solanas, 1992).

El método Monte Carlo requiere que se especifique una función de distribución de probabilidad  $F(X)$ , cuyos parámetros sean conocidos o estimados, y a partir de ella, generar números aleatorios que sean realizaciones de variables aleatorias que sigan dicha distribución.

Es decir, el Bootstrap paramétrico utiliza ampliamente la simulación Monte Carlo, ya que se generan los datos de una distribución en base a parámetros fijados de antemano.

El primer paso consiste en especificar la función de distribución  $F(X)$  de la cual se generará una cantidad  $n$  de datos. A continuación, se genera una distribución Uniforme Estándar, y se obtiene un dato  $\alpha$  de esta distribución, es decir:

$$\alpha_i \sim U(0, 1).$$

Supongamos que  $X$  es una variable aleatoria con distribución  $F$ , obtenemos una realización  $x_i$  de la variable aleatoria  $X$ .

Obtenemos el valor de  $x_i$  al representar a este valor como la preimagen de la función de distribución  $F$  para la cuál se tiene una probabilidad igual a  $\delta_i$ , es decir, obtenemos:

$$x_i = F^{-1}(\alpha_i)$$

Estos pasos se repiten hasta obtener la cantidad de datos requeridos.

### 3.4 La distribución bootstrap

El propósito fundamental del bootstrap no paramétrico es obtener una estimación de la distribución bootstrap del estadístico de interés, debido a que la misma puede brindarnos información sobre la distribución verdadera del estadístico. Esta semejanza se puede establecer en base al principio *plug-in*.

La idea del principio *plug-in* es que “para estimar un parámetro poblacional, se utiliza el estadístico correspondiente para cada muestra” (Hesterberg, 2003). Es decir, es posible usar la distribución bootstrap de un estadístico para estimar el valor de un parámetro poblacional, razón por la cual la distribución bootstrap es tan importante.

Las distribuciones bootstrap tienen “la misma forma y dispersión que la distribución muestral del estadístico, pero centradas en el valor del estadístico en la muestra original” (Hesterberg, 2003), por lo que es muy importante que la muestra original sea representativa de la población.

La distribución bootstrap está centrada en el valor del estadístico en la muestra original, la distribución Bootstrap no brinda información sobre el centro de la distribución muestral del estadístico, pero nos brinda importante información sobre el sesgo, puesto que “pese a que la distribución Bootstrap y Muestral no tienen el mismo centro, tienen el mismo sesgo” (Hesterberg, 2003).

Si bien la estimación *plug-in* de un estadístico se forma a base de la distribución empírica de los datos, existen otras metodologías con el propósito de determinar la distribución  $F$ . En general método Bootstrap se propone estimar  $F$  a través de  $\hat{F}$ , para lo cual consideraremos dos posibilidades: el bootstrap paramétrico y no paramétrico.

### 3.5 El método bootstrap para determinar diferencias entre grupos

Las pruebas hechas para determinar si existe diferencias entre grupos se contrastan en base a la hipótesis nula de que no existen diferencias entre tales grupos. Por tal razón “no es posible realizar remuestreo de la muestra original, es necesario crear una distribución centrada en el valor del parámetro establecido por la hipótesis nula” (Hesterberg, 2003).

En base a la información obtenida, se compara el valor observado del parámetro en la muestra con respecto a la distribución Bootstrap que se esperaría obtener al ser generada bajo la hipótesis nula.

En este aspecto, el Bootstrap se aplica en estudios donde no se puede determinar si ha existido un proceso de muestreo aleatorio para la conformación de los grupos. Este es el caso de los Estudios Interlaboratorio donde las mediciones se recopilan por separado en cada laboratorio, y el investigador no tiene influencia sobre la forma en que se presentan las mediciones.

Los pasos por seguir para determinar mediante Bootstrap si existe diferencia entre dos o más grupos es la siguiente:

1. Se mezclan las observaciones de los  $k$  grupos en un solo conjunto, conocido como el conjunto *pool* de los datos.
2. Se aplica el muestreo con reemplazamiento sobre este conjunto para generar una remuestra del mismo tamaño.
3. Se vuelven a formar  $k$  grupos del mismo tamaño que los originales, asignando de forma aleatoria las observaciones obtenidas en el paso anterior.
4. Se calcula el valor del estadístico de interés en cada nuevo grupo.
5. Se repite este proceso una gran cantidad de veces (por ejemplo 500, 1000, 10000) para poder formar la distribución Bootstrap. Esta es la distribución del estadístico producida debido

al muestreo aleatorio de los datos de los grupos y asumiendo que todas las observaciones provienen de la misma población.

El método bootstrap para establecer diferencia entre grupos se aplicará en la presente investigación, para lo cual, cada laboratorio se considerará como un grupo, y los estadísticos que calcularemos en cada grupo serán los estadísticos  $h$  y  $k$  de Mandel.

El objetivo será determinar si la diferencias entre los laboratorios se deben a un determinado factor: error en la aplicación del método de medición o error en la instrumentación empleada, o se debe a la “variabilidad que se puede esperar debido a un proceso de muestreo aleatorio” (Amat, 2016).

## 4 Metodología

### 4.1 Introducción

En este capítulo desarrollaremos una metodología que no dependa del supuesto de normalidad, el cual es necesario para realizar Estudios Interlaboratorio, en concreto, nuestro objetivo es desarrollar un método que calcule los valores críticos de los estadísticos  $h$  y  $k$  de Mandel en base a la distribución bootstrap de estos estadísticos.

Las normas ASTM E-691 e ISO 5725 proponen calcular los valores críticos del estadístico  $h$  de Mandel mediante una distribución  $t$  de Student, y por otra parte, determinar el valor crítico del estadístico  $k$  de Mandel mediante una distribución  $F$ , las cuáles son distribuciones que requieren del cumplimiento de ciertas condiciones, como el principio de normalidad.

Tomando en cuenta que el análisis requerido en un Estudio Interlaboratorio es un ANOVA unidireccional de un solo factor, las condiciones que deben cumplirse son las clásicas de este modelo, es decir, los datos de todos los grupos (laboratorios) deben provenir de la misma población (la distribución normal), y la media y la desviación estándar entre los grupos debe ser similar. Sin embargo, en la práctica, no siempre se verifica el cumplimiento de estas condiciones antes de iniciar un Estudio Interlaboratorio.

Nuestra propuesta es que los puntos críticos para los estadísticos  $h$  y  $k$  sean calculados mediante la distribución bootstrap de estos estadísticos, la cual se obtendrá mediante un algoritmo que haga uso del método bootstrap no paramétrico. Calcularemos las distribuciones bootstrap de los estadísticos  $h$  y  $k$  de Mandel, de manera que las mismas se asemejen a las distribuciones reales de estos estadísticos.

## 4.2 Descripción de la metodología propuesta

En esta sección presentamos la metodología desarrollada para obtener los valores críticos de los estadísticos  $h$  y  $k$  de Mandel mediante el método bootstrap. La metodología propuesta se fundamenta en un algoritmo que fue implementado para estimar los valores críticos correspondientes a la distribución bootstrap de los estadísticos  $h$  y  $k$ .

Como ya se dijo anteriormente, un Estudio Interlaboratorio se desarrolla bajo la hipótesis nula de que todas las observaciones de los diferentes laboratorios vienen de la misma población, es decir, la distribución normal. Lo que equivale a decir, en el contexto de un modelo ANOVA, que las medias y las desviaciones estándar de todos los laboratorios son aproximadamente iguales para cada nivel de medición.

Esto nos permitirá plantear las hipótesis de reproducibilidad y repetibilidad a ser contrastadas durante esta investigación. La hipótesis nula de reproducibilidad, establece que no existen diferencias entre las medias de los laboratorios, mientras que la hipótesis nula de repetibilidad, establece que no existen diferencias entre las desviaciones estándar de los laboratorios.

## 4.3 Hipótesis de reproducibilidad

La hipótesis nula de reproducibilidad que proponemos establece que no hay diferencia entre las medias de los laboratorios, es decir:

$$H_o = \mu_1 = \mu_2 = \dots = \mu_l \quad (11)$$

Donde  $L = 1, 2, 3, \dots, l$  es el número de laboratorios, y  $\mu_l$  es la media de cada laboratorio.

Para constrar la hipótesis nula de reproducibilidad dada en (11) utilizaremos el estadístico  $h$  de Mandel definido en (2) mediante bootstrap no paramétrico.



#### 4.4 Hipótesis de repetibilidad

La hipótesis nula de repetibilidad que proponemos establece que no hay diferencia entre las desviaciones estándar de los laboratorios, es decir:

$$H_o = \sigma_1 = \sigma_2 = \dots = \sigma_l \quad (12)$$

Donde  $L = 1, 2, 3, \dots, l$  es el número de laboratorios, y  $\sigma_l$  es la desviación estándar de cada laboratorio.

Constrastamos la hipótesis de repetibilidad dada en (12) mediante el estadístico  $k$  que fue definido en (4) calculado mediante técnicas de bootstrap no paramétrico.

#### 4.5 Algoritmo para la detección de datos atípicos

El algoritmo propuesto que hace uso del bootstrap no paramétrico pretende reproducir la distribución de los estadísticos  $h$  y  $k$  de Mandel bajo las hipótesis nulas de reproducibilidad (11) y repetibilidad (12). Suponiendo fijo un nivel de significancia  $\alpha = 0.010$ , el algoritmo consta de los siguientes pasos:

1. Se agrupan los datos provenientes de los diferentes laboratorios en un solo conjunto de tamaño  $L \times n$ .
2. Se elabora un diagrama de caja de los datos, luego se separan los datos que están fuera del rango intercuantíl.
3. Se realiza muestreo con reemplazamiento con los datos que no fueron separados para generar una remuestra del mismo tamaño del conjunto en el que se mezclaron los datos originalmente.
4. Se asignan de manera aleatoria las observaciones de la remuestra en grupos de igual tamaño a los laboratorios originales, es decir, asignamos  $n$  datos a  $L$  laboratorios.
5. Dentro de la remuestra se calcula el valor del estadístico de interés ( $h$  o  $k$ ) para cada laboratorio.

6. Se repiten los pasos 3, 4, 5, 6 un gran número de veces, por ejemplo,  $B = 500$ . Con esto se obtiene la distribución Bootstrap del estadístico.
7. **Para el estadístico  $h$ :** Se determinan los valores críticos del estadístico  $h$ , calculando los cuantiles de la distribución bootstrap del estadístico  $h$  para los niveles  $\frac{\alpha}{2}$  y  $1 - \frac{\alpha}{2}$ .
8. Se calculan los estadísticos  $h$  en la muestra original, si un valor de  $h$  es mayor que el valor absoluto del valor crítico del estadístico  $h$ , el laboratorio correspondiente se considera como atípico.
9. **Para el estadístico  $k$ :** Se determina el valor crítico del estadístico  $k$ , calculando el cuantil de la distribución bootstrap del estadístico  $k$  para el nivel  $1 - \alpha$ .
10. Se calculan los estadísticos  $k$  en la muestra original, si un valor de  $k$  es mayor que valor crítico del estadístico, el laboratorio correspondiente se considera como atípico.

Esta metodología fue aplicada en el análisis de un Estudio Interlaboratorio que se desarrolló como parte de esta investigación. El Estudio Interlaboratorio se centró en analizar el método estándar de Biometría Hemática Mediante Técnica Computarizada, el cual fue aplicado por un grupo de laboratorios clínicos para determinar el nivel de plaquetas, hemoglobina, globulos rojos, globulos bancos, entre otros, en una muestra sanguínea.

Sin embargo, para poder validar el uso de la propuesta es necesario evaluarla en varios escenarios, por lo que desarrollamos un estudio de simulación en el cuál aplicamos la metodología propuesta con datos provenientes de distintas distribuciones con diferente cantidad de datos atípicos.

#### 4.6 Validación de la metodología propuesta

La metodología propuesta para la detección de datos atípicos fue evaluada mediante un estudio de simulación, en el cual, consideraremos escenarios que se asemejen a los que se podrían presentar

en un Estudio Interlaboratorio en el caso de contar con diferente cantidad de datos atípicos.

Para lograr este objetivo hemos elegido tres distribuciones distintas que simularon la presencia de datos atípicos. Estas distribuciones fueron la distribución normal, la cual es simétrica con poco peso en las colas. La distribución de Laplace, la cual es una distribución simétrica pero con mayor peso en las colas. Y una distribución sesgada a la derecha que fue obtenida a partir de la expresión  $|X| - \sqrt{\frac{2}{\pi}}$ , donde  $X$  es una variable aleatoria que sigue una distribución normal, es decir,  $X \sim N(\mu, \sigma^2)$ .

Para la generar datos de la distribución de Laplace utilizamos el paquete *rmutil* version 1.1.3 (Lyndsey and Swihart, 2019) disponible en el repositorio de **R**. Para calcular los valores críticos de los estadísticos  $h$  y  $k$  de Mandel de manera tradicional, es decir según lo dispuesto en la norma ISO 5725, utilizamos el paquete *metRology* version 0.9-28-1 (Ellison, 2018) disponibles en disponibles en <https://cran.r-project.org>.

La primera parte de este estudio de simulación consistió en evaluar el cumplimiento de la hipótesis de reproducibilidad (11) para las distintas distribuciones consideradas y para diferentes tamaños de muestras, para ello variamos la media de cada distribución, manteniendo constante el valor de la desviación estándar. Este procedimiento lo repetimos para cada una de las tres distribuciones consideradas.

Luego evaluamos el cumplimiento de la hipótesis de repetibilidad (12) para las distintas distribuciones consideradas, para ello variamos el parámetro correspondiente a la desviación estándar y mantuvimos fijo el valor de la media.

A continuación, se presentan las distribuciones consideradas para este estudio de simulación, y los escenarios de consistencia e inconsistencia de datos que hemos tomado en cuenta.

#### 4.6.1 La distribución normal

Esta distribución representa a las distribuciones simétricas y con poco peso en sus colas. La distribución que recreó el escenario de consistencia de los datos fue la distribución normal estándar, ( $\mu = 0, sd = 1$ ).

#### 4.6.2 La distribución de Laplace

Esta distribución representa a las distribuciones simétricas con mayor peso en las colas. La distribución que simula el escenario de consistencia de datos es una distribución de Laplace con parámetro de localización ( $\mu = 0$ ), y parámetro de escala  $b = 1$

#### 4.6.3 La distribución sesgada

Para obtener una distribución sesgada se ha procedido a realizar la siguiente transformación a una variable aleatoria  $X$  que sigue una distribución normal, es decir: dad  $X \sim N(\mu, \sigma^2)$ , la transformación considerada fue:

$$|X| - \sqrt{\frac{2}{\pi}}$$

Las distribuciones generadas presentan un sesgo hacia la derecha. De esta forma, estas distribuciones representarán a las distribuciones sesgadas y asimétricas.

### 4.7 Descripción del estudio de simulación

Para determinar la potencia de la prueba hemos dispuesto que los datos de  $(L - 1)$  laboratorios sean generados bajo una condición de consistencia de datos, mientras que un laboratorio ( $m = 1$ ) sea generado bajo condiciones de inconsistencia, es decir, el último laboratorio de la simulación sufrirá variaciones en su media o desviación estándar para la distribución en estudio. Se ha simulado, por

lo tanto, la existencia de  $(L + m)$  laboratorios, los cuáles realizarán  $mc = 1000$  estudios interlaboratorio sobre un solo material. Dentro de cada Estudio Interlaboratorio se realizarán  $B = 500$  remuestras de los datos de los  $(L + m)$  laboratorios.

De manera concreta incluiremos  $L$  laboratorios de tamaños  $\{5, 10\}$ , y dentro de cada laboratorio hemos generado  $n$  observaciones o mediciones, de tamaños  $\{3, 6\}$ , manteniendo fijo el nivel de significancia para  $\alpha = 0.01$ . Esperamos que como resultado del estudio de simulación, al aumentar el tamaño  $n$  de la muestra, aumente la capacidad del algoritmo propuesto para detectar datos atípicos (potencia de la prueba), de igual forma, esperamos que al aumentar la variación en la media o la desviación estándar, aumente la capacidad del algoritmo para detectar datos atípicos.

El algoritmo descrito en la sección anterior, ha sido aplicado en el estudio de simulación que se propone como parte del presente trabajo. Este estudio de simulación consistirá en la aplicación del método bootstrap no paramétrico.

En la siguiente sección presentamos las variaciones que realizamos a la media y a la varianza para generar el laboratorio inconsistente dentro de cada simulación.

#### 4.7.1 Incumplimiento de la hipótesis de reproducibilidad

Sean  $\Delta_\mu$  las variaciones que realizamos para la media en cada una de las distribuciones, se ha procedido a variar la media en un solo laboratorio, considerado como el laboratorio atípico. Para el laboratorio atípico se ha considerado los siguientes cambios en la media:

$$\Delta_\mu \in \{-3, -2, -1, 0, 1, 2, 3\}.$$

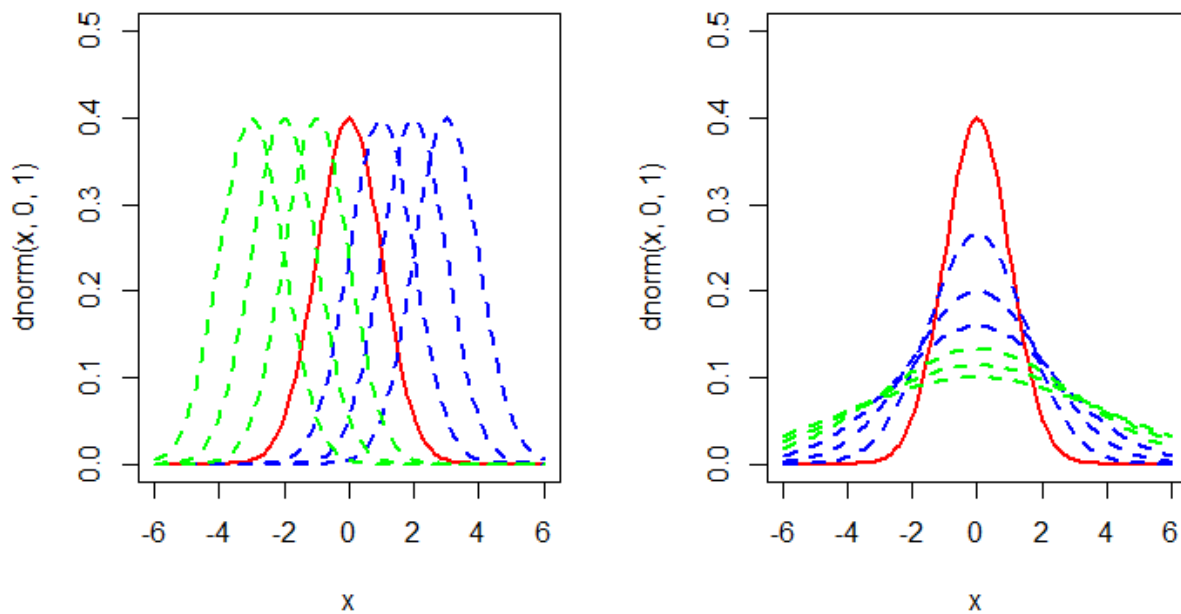
#### 4.7.2 Incumplimiento de la hipótesis de repetibilidad

Sean  $\Delta_\sigma$  las variaciones que se realizaron a la desviación estándar del laboratorio asignado como atípico, se procederá a variar la desviación estándar de este laboratorio para producir cambios en la varianza de la distribución normal. Para la distribución de Laplace hemos procedido a variar el parámetro de escala  $b$ .

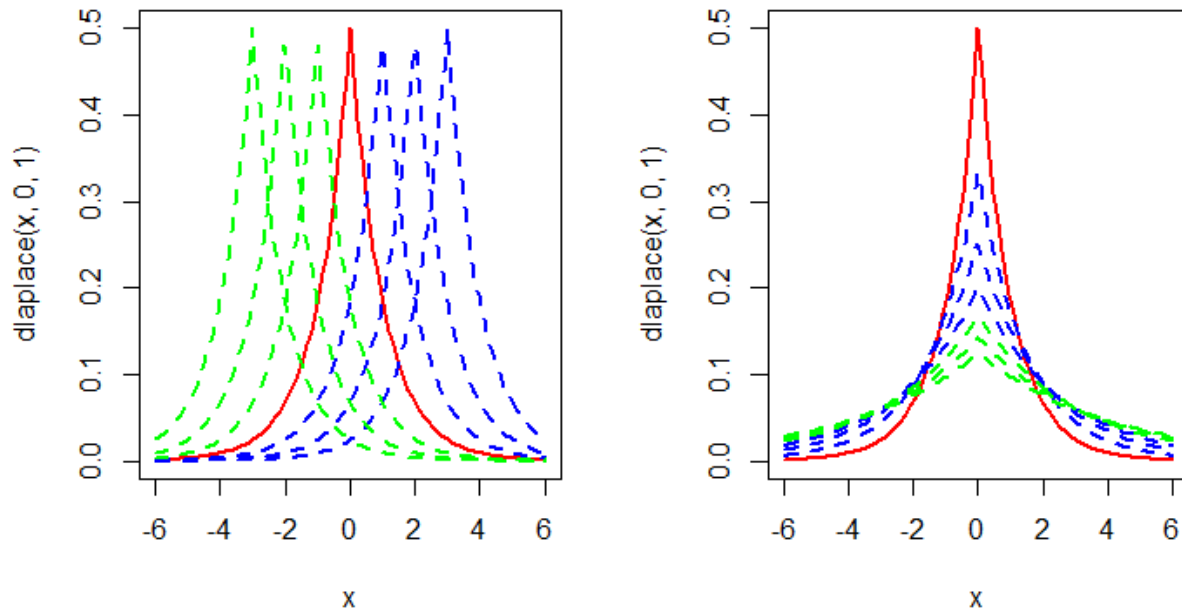
A partir de los escenarios de consistencia creamos un laboratorio que sea inconsistente, para el cual hemos considerado los siguientes cambios en la desviación estándar:

$$\Delta_\sigma \in \{1, 1.5, 2, 2.5, 3, 3.5, 4\}.$$

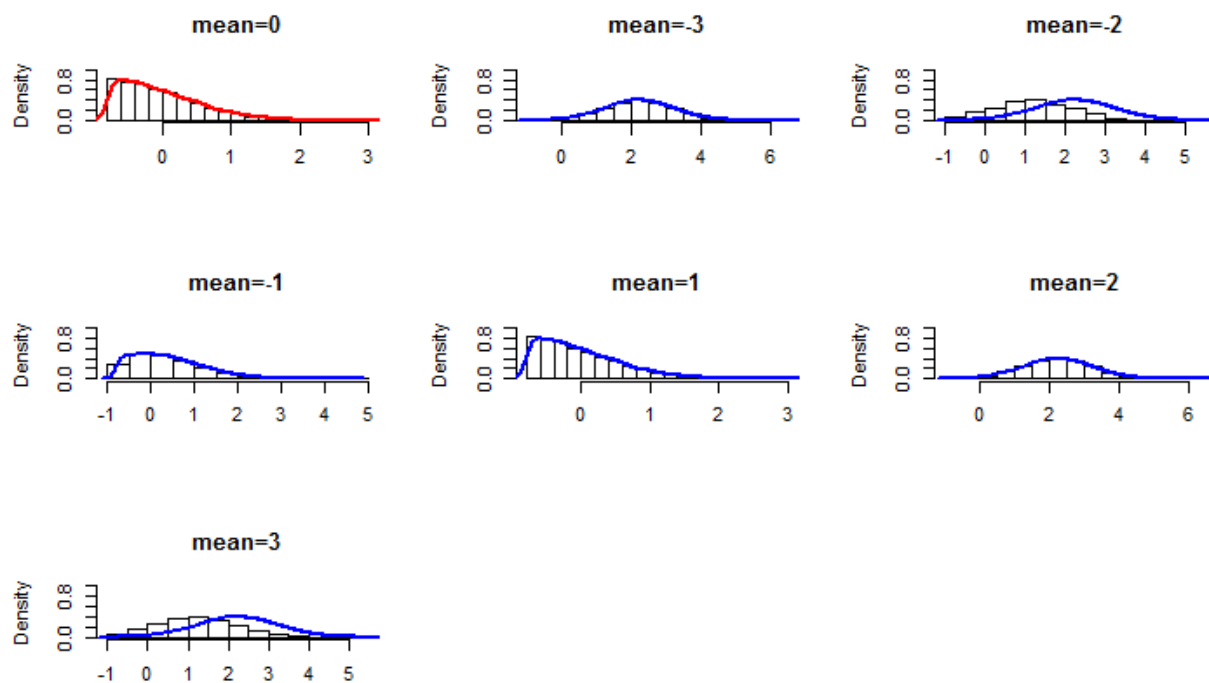
A continuación presentamos gráficamente las hipótesis de reproducibilidad y repetibilidad que se manejaron durante el estudio:



**Figura 1:** Variaciones en la media y la desviación estándar para la distribución normal.

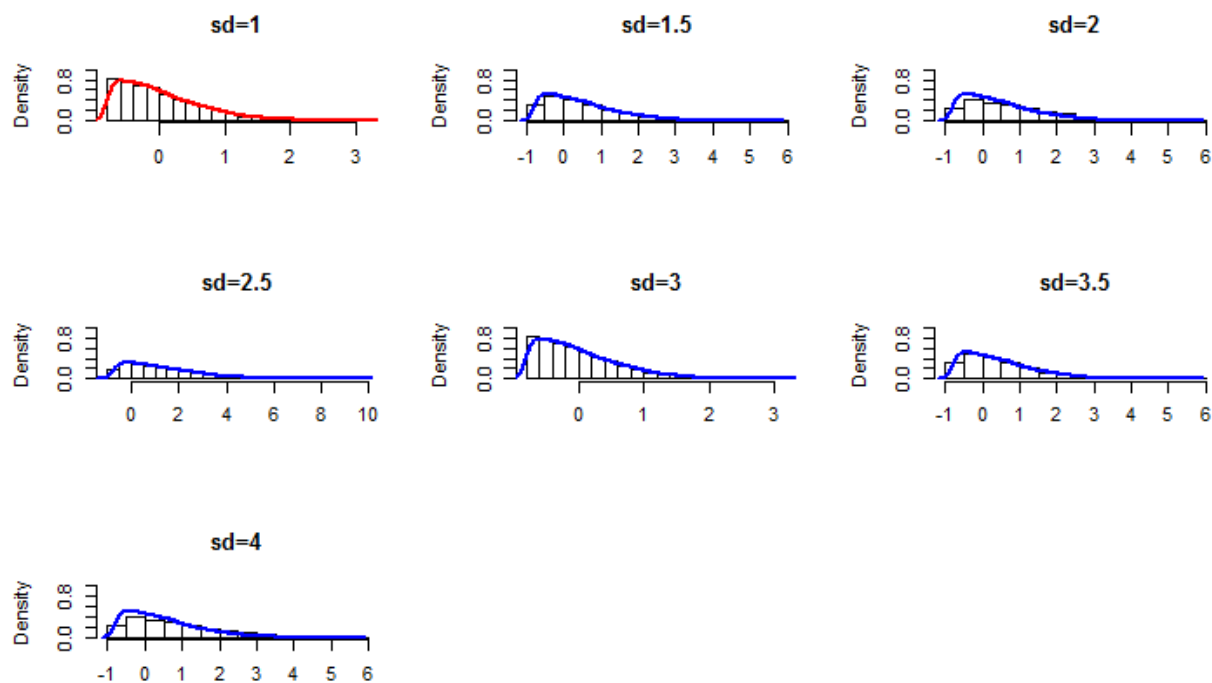


**Figura 2:** Variaciones en la media y el factor de escala para la distribución de Laplace.



*Figura 3:* Variaciones en la media para la distribución sesgada.





*Figura 4:* Variaciones en la desviación estándar para la distribución sesgada.

## 5 Estudio de Simulación

### 5.1 Introducción

En la primera sección de este capítulo presentamos los resultados obtenidos en el estudio de simulación cuando se ha incumplido con las hipótesis de reproducibilidad y de repetibilidad, para las tres distribuciones consideradas: la distribución normal, la distribución de Laplace, y la distribución sesgada descrita en el capítulo anterior.

Fueron considerados dos escenarios para evaluar la proporción de rechazo de la hipótesis nula, es decir, la proporción de veces que un laboratorio es detectado como atípico cuando sus datos provienen de una distribución distinta al del resto de laboratorios.

Vale recalcar que la metodología propuesta fue diseñada para detectar por separado cambios en la media y en la desviación estándar de los resultados de las mediciones de los laboratorios.

### 5.2 Proporción de rechazo de la hipótesis nula de reproducibilidad

Para evaluar la proporción de rechazo de la hipótesis nula de reproducibilidad, ( $L = 9$ ) laboratorios fueron simulados bajo la hipótesis nula (los datos provienen de la distribución), mientras que un laboratorio ( $m = 1$ ) fue generado con los valores alternativos para la media que se indicaron en el anterior capítulo.

Se aplicó el algoritmo propuesto en el capítulo anterior para estimar la distribución Bootstrap del estadístico  $h$ , y luego procedimos a calcular los valores críticos del estadístico, considerando un valor de significancia constante  $\alpha = 0.01$ .

A continuación, presentamos los resultados de las simulaciones realizadas para detectar la proporción de rechazo de la hipótesis nula de reproducibilidad para las diferentes distribuciones.

### 5.2.1 Incumplimiento de la hipótesis nula de reproducibilidad para la distribución normal

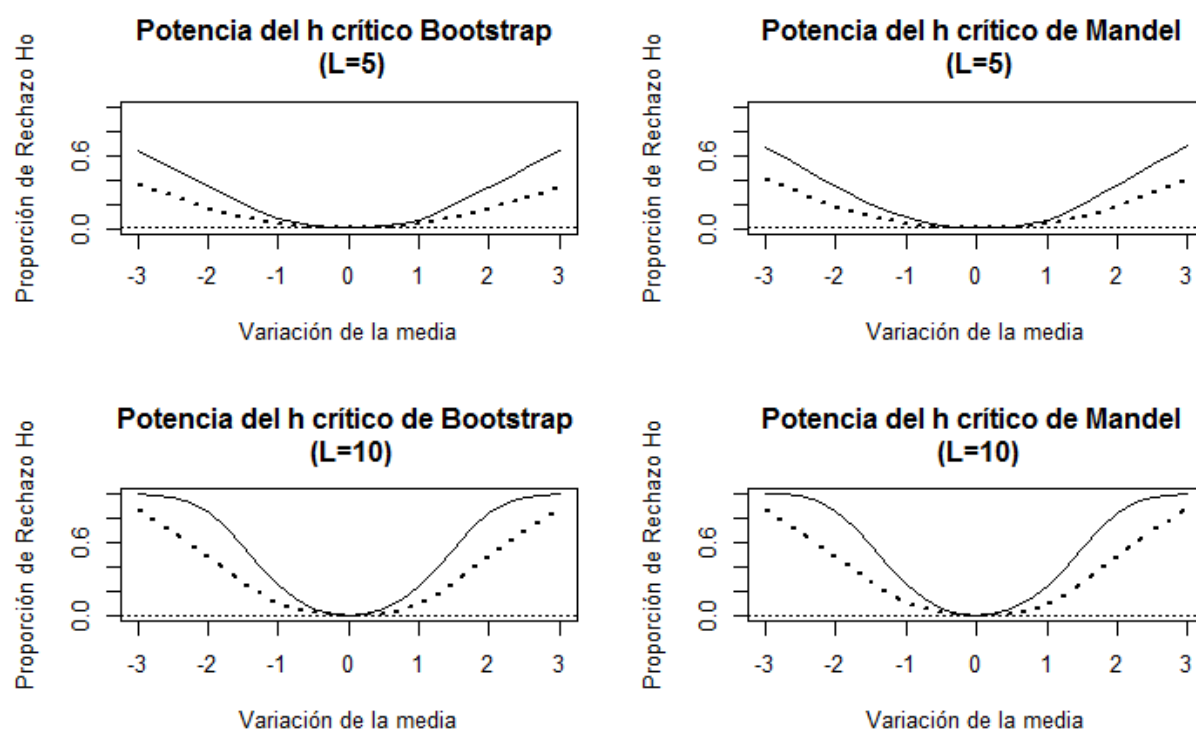
**Tabla 1:**

*Potencia del estadístico  $h$  para la distribución normal*

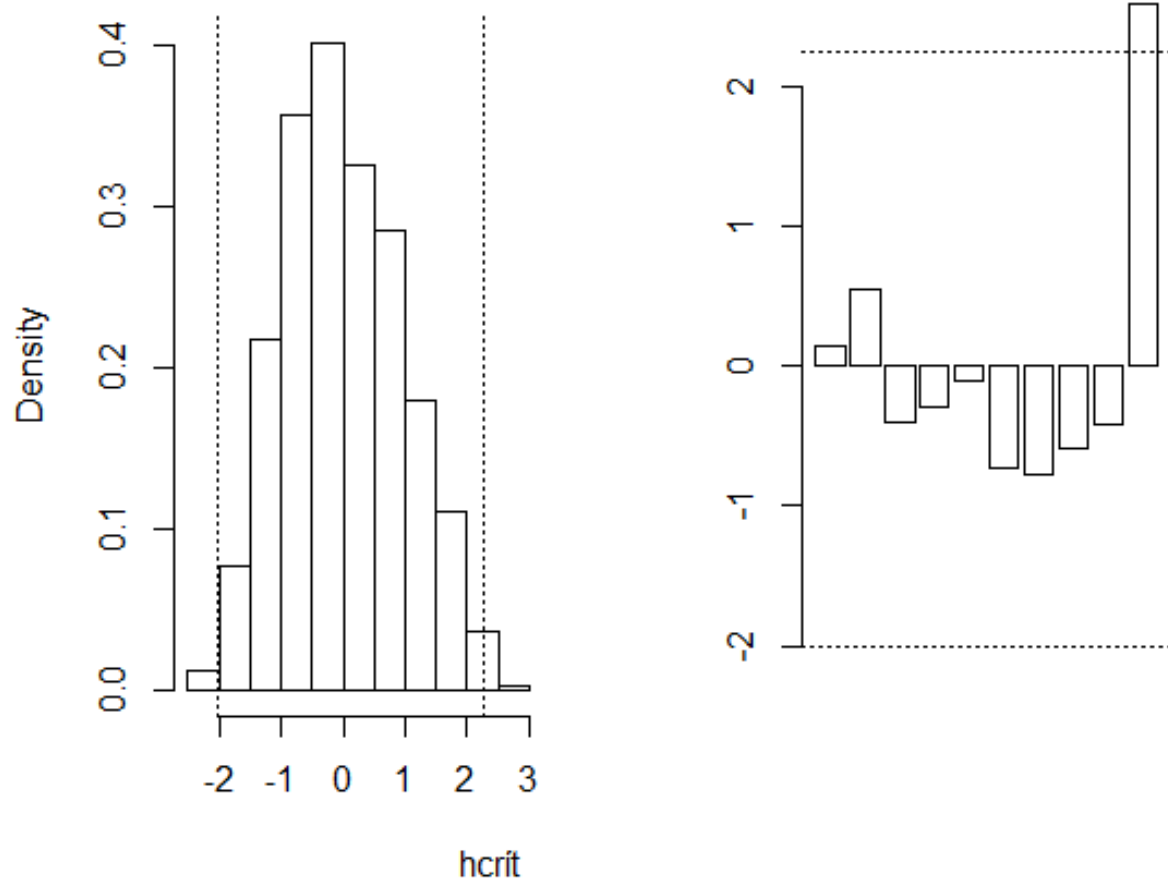
$L$	$n$	Estadístico $h$ Bootstrap ( $\Delta_\mu$ )						Estadístico $h$ univariante ( $\Delta_\mu$ )							
		(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)	(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)
5	3	0.367	0.170	0.045	0.009	0.044	0.167	0.353	0.416	0.180	0.042	0.009	0.043	0.188	0.408
	6	0.634	0.360	0.090	0.009	0.077	0.343	0.640	0.673	0.350	0.092	0.010	0.076	0.358	0.676
10	3	0.869	0.495	0.108	0.012	0.105	0.496	0.889	0.889	0.508	0.107	0.008	0.094	0.499	0.919
	6	0.996	0.856	0.258	0.008	0.241	0.841	0.996	0.997	0.860	0.259	0.009	0.242	0.849	0.999

En la tabla (1) se han registrado las proporciones de rechazo de la hipótesis nula cuando los datos provienen de una distribución normal comparando los resultados obtenidos por el método bootstrap y el método tradicional, estos resultados se hallan resumidos en la figura (5). En la figura podemos comparar los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda). Puede verse que conforme aumenta el tamaño de la variación de la media, aumenta la potencia del estadístico  $h$ ; de igual forma, conforme aumenta el tamaño  $n$  de las muestras (la línea segmentada cuando  $n = 3$  y la línea continua cuando  $n = 6$ ), aumenta la potencia de la prueba.

Podemos notar que el método tradicional y el método bootstrap propuesto obtienen resultados similares, esto se explica porque el método tradicional asume que los datos siguen la distribución normal, y por lo tanto, los resultados obtenidos con el método bootstrap van a ser muy similares. En la figura (6) podemos observar la distribución muestral del estadístico para una de las iteraciones del algoritmo bootstrap, puede verse como el último laboratorio ha sido detectado como atípico



**Figura 5:** Curvas de potencia del estadístico  $h$  para la distribución normal.



**Figura 6:** La distribución bootstrap y los valores  $h$  críticos para la distribución normal.

## 5.2.2 Incumplimiento de la hipótesis nula de reproducibilidad para la distribución de Laplace

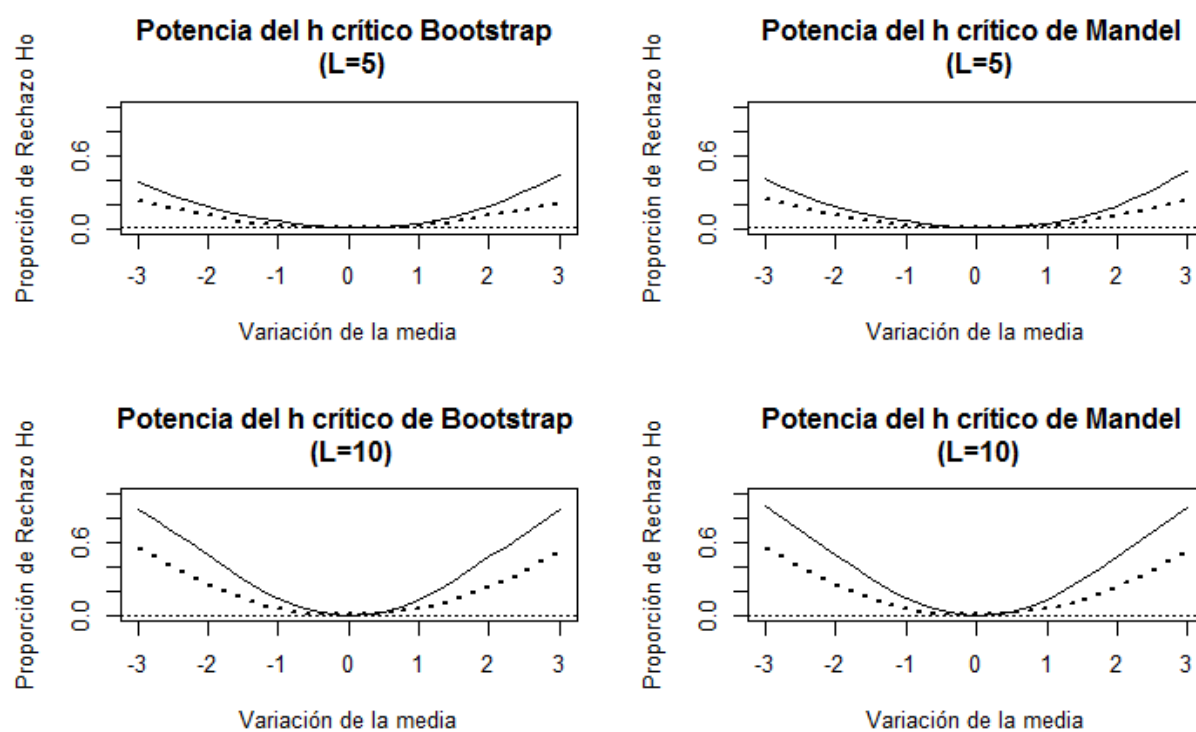
**Tabla 2:**

*Potencia del estadístico  $h$  para la distribución de Laplace*

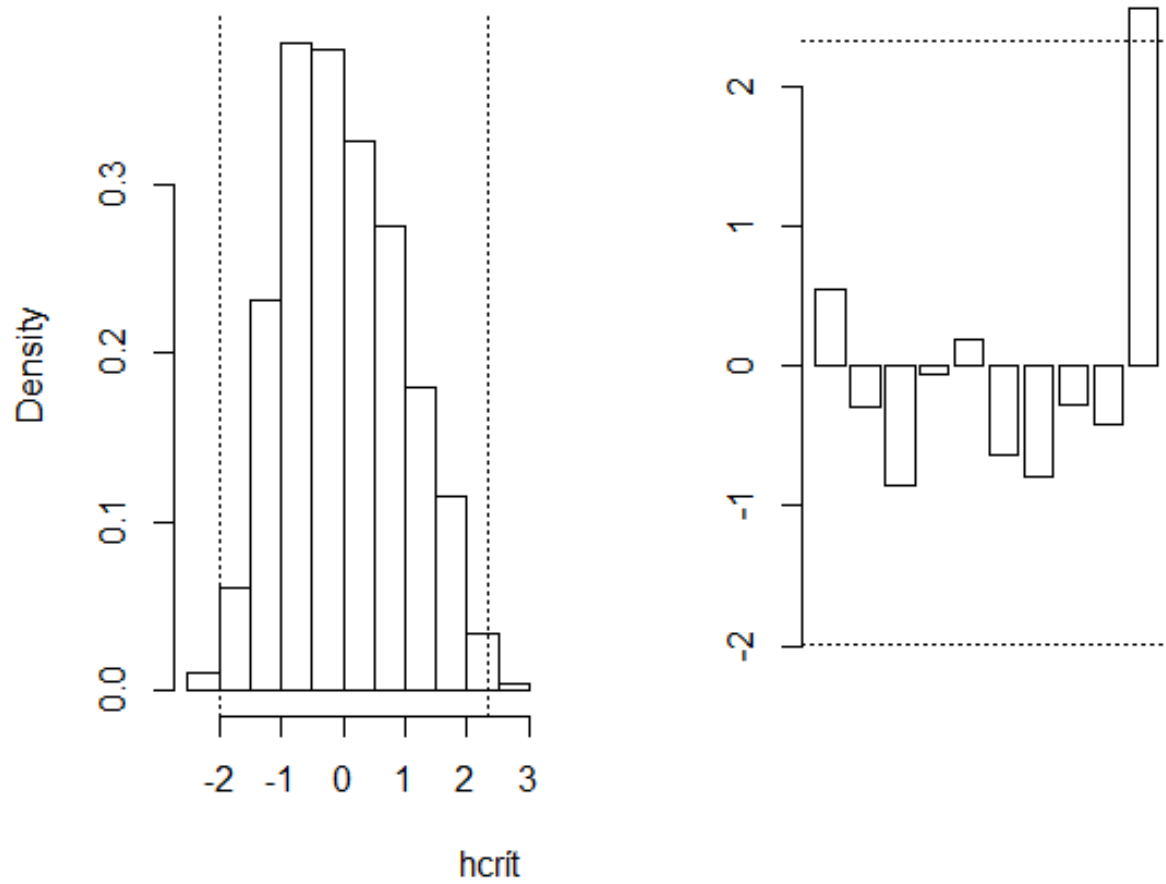
$L$	$n$	Estadístico $h$ Bootstrap ( $\Delta_\mu$ )						Estadístico $h$ univariante ( $\Delta_\mu$ )							
		(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)	(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)
5	3	0.233	0.106	0.029	0.012	0.028	0.108	0.219	0.247	0.111	0.029	0.012	0.029	0.109	0.240
	6	0.386	0.178	0.067	0.010	0.043	0.189	0.442	0.412	0.177	0.066	0.010	0.041	0.188	0.470
10	3	0.560	0.265	0.061	0.019	0.067	0.244	0.534	0.605	0.268	0.053	0.015	0.067	0.246	0.573
	6	0.877	0.527	0.120	0.011	0.119	0.503	0.873	0.899	0.529	0.117	0.011	0.113	0.522	0.893

En la tabla (2) se han registrado las proporciones de rechazo de la hipótesis nula cuando los datos siguen la distribución de Laplace comparando el método bootstrap y el método tradicional, esta información se encuentra resumida en la figura (7). En la misma podemos comparar los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda), se puede observar que al aumentar el tamaño de la variación de la media, aumenta la potencia del estadístico  $h$ , de igual manera, conforme aumenta el tamaño  $n$  de las muestras (la línea segmentada cuando  $n = 3$  y la línea continua cuando  $n = 6$ ) aumenta la proporción de rechazo de la hipótesis nula.

En general ambos métodos obtienen resultados similares, esto se explica porque la distribución de Laplace es simétrica y centrada en el valor de la media, por lo tanto, la distribución muestral bootstrap va a reproducir estas características. En la figura (8) podemos observar la distribución muestral del estadístico para una de las iteraciones del algoritmo bootstrap, puede verse como el último laboratorio ha sido detectado como atípico.



**Figura 7:** Curvas de potencia del estadístico  $h$  para la distribución de Laplace.



**Figura 8:** La distribución bootstrap y valores  $h$  críticos para la distribución de Laplace.



### 5.2.3 Incumplimiento de la hipótesis nula de reproducibilidad para la distribución sesgada

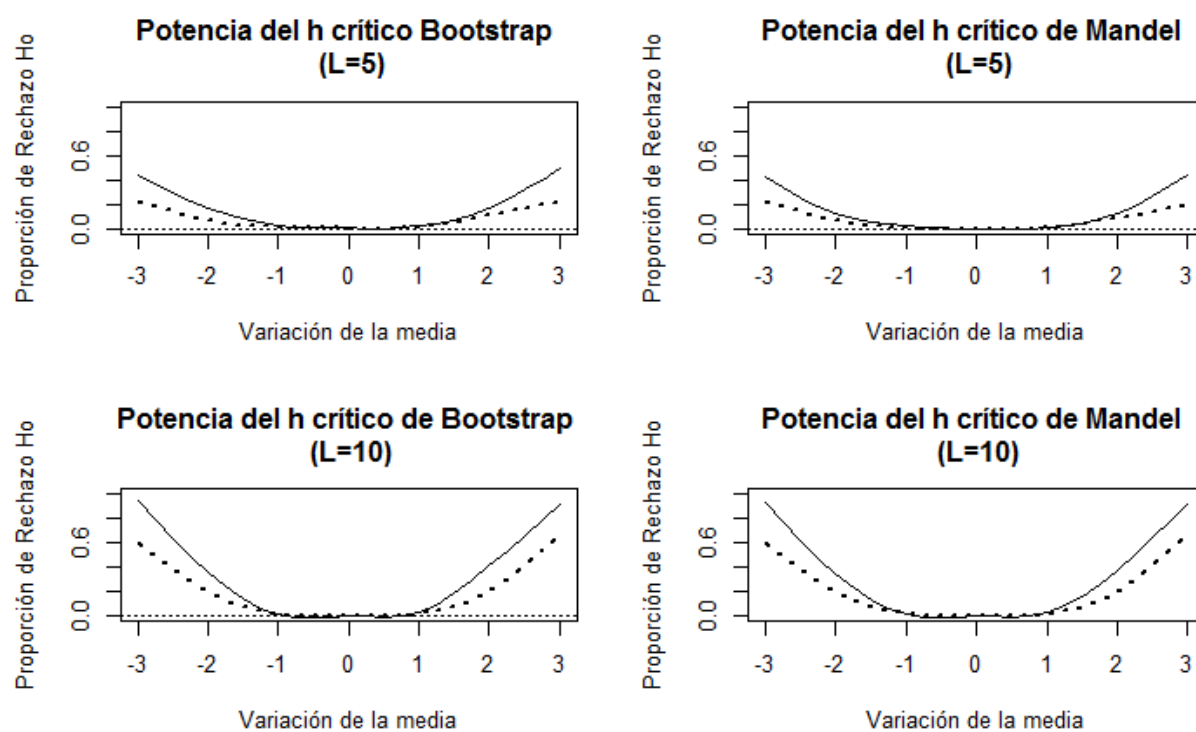
**Tabla 3:**

*Potencia del estadístico  $h$  para la distribución distribución sesgada.*

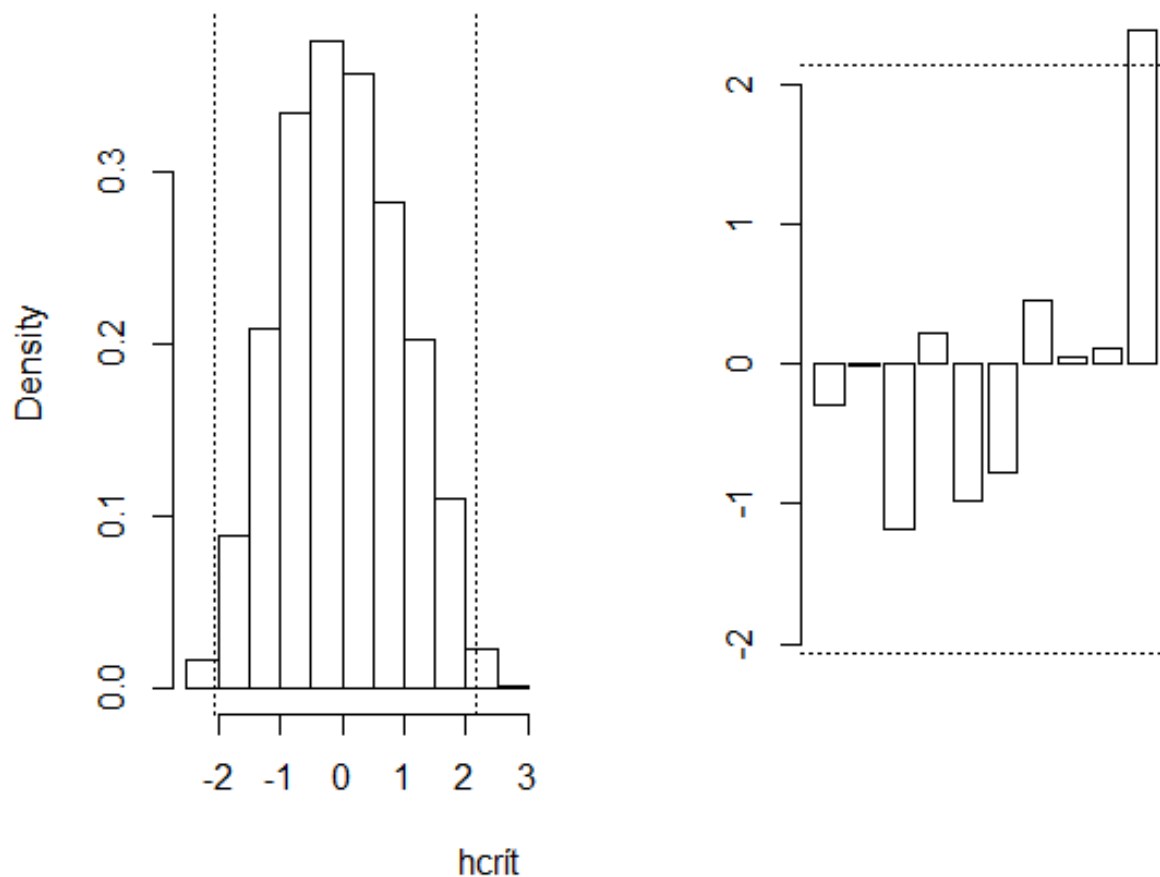
$L$	$n$	Estadístico $h$ Bootstrap ( $\Delta_\mu$ )							Estadístico $h$ univariante ( $\Delta_\mu$ )						
		(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)	(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)
5	3	0.232	0.066	0.018	0.004	0.020	0.116	0.216	0.236	0.058	0.010	0.002	0.010	0.092	0.202
	6	0.440	0.170	0.028	0.006	0.022	0.174	0.490	0.424	0.124	0.026	0.002	0.012	0.126	0.440
10	3	0.608	0.202	0.018	0.002	0.018	0.210	0.668	0.596	0.182	0.010	0.002	0.012	0.190	0.656
	6	0.940	0.364	0.016	0.004	0.036	0.414	0.918	0.934	0.344	0.014	0.002	0.030	0.382	0.916

En la tabla (3) se han registrado las proporciones de rechazo de la hipótesis nula cuando los datos provienen de una distribución sesgada a la derecha comparando el método bootstrap con el método tradicional, estos resultados se hallan resumidos en la figura (9). En la misma puede compararse los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda), podemos ver que conforme aumenta el tamaño  $n$  de la muestra (línea segmentada cuando  $n = 3$  y línea continua cuando  $n = 6$ ) aumenta la potencia de la prueba, en especial, para tamaños pequeño de muestras, y un distribución sesgada, el método bootstrap detecta en mayor proporción los datos atípicos que el método tradicional.

La distribución bootstrap del estadístico  $h$  puede verse en la figura (10), puede verse como el último laboratorio ha sido detectado como atípico.



**Figura 9:** Curvas de potencia del estadístico  $h$  para la distribución sesgada.



*Figura 10:* La distribución bootstrap y los valores  $h$  críticos para la distribución sesgada.

### 5.3 Proporción de rechazo de la hipótesis nula de repetibilidad

Para evaluar la potencia del estadístico  $k$  para detectar variaciones en la desviación estándar hemos consideramos ( $L = 9$ ) laboratorios que fueron simulados bajo la hipótesis nula, mientras que un laboratorio ( $m = 1$ ) fue generado con los valores alternativos para la desviación estándar que se indicaron en el anterior capítulo.

A continuación presentamos los resultados del estudio de simulación para las distintas distribuciones.

### 5.3.1 Incumplimiento de la hipótesis nula de repetibilidad para la distribución normal

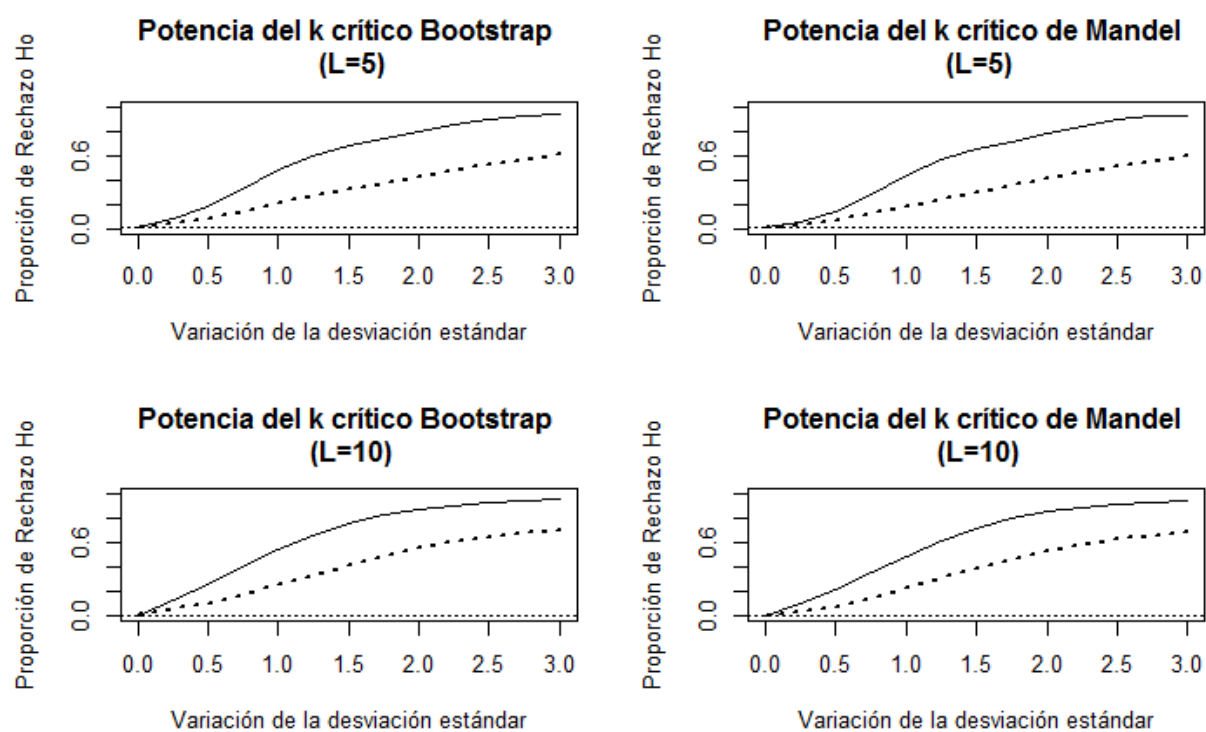
**Tabla 4:**

*Potencia del estadístico  $k$  para la distribución normal*

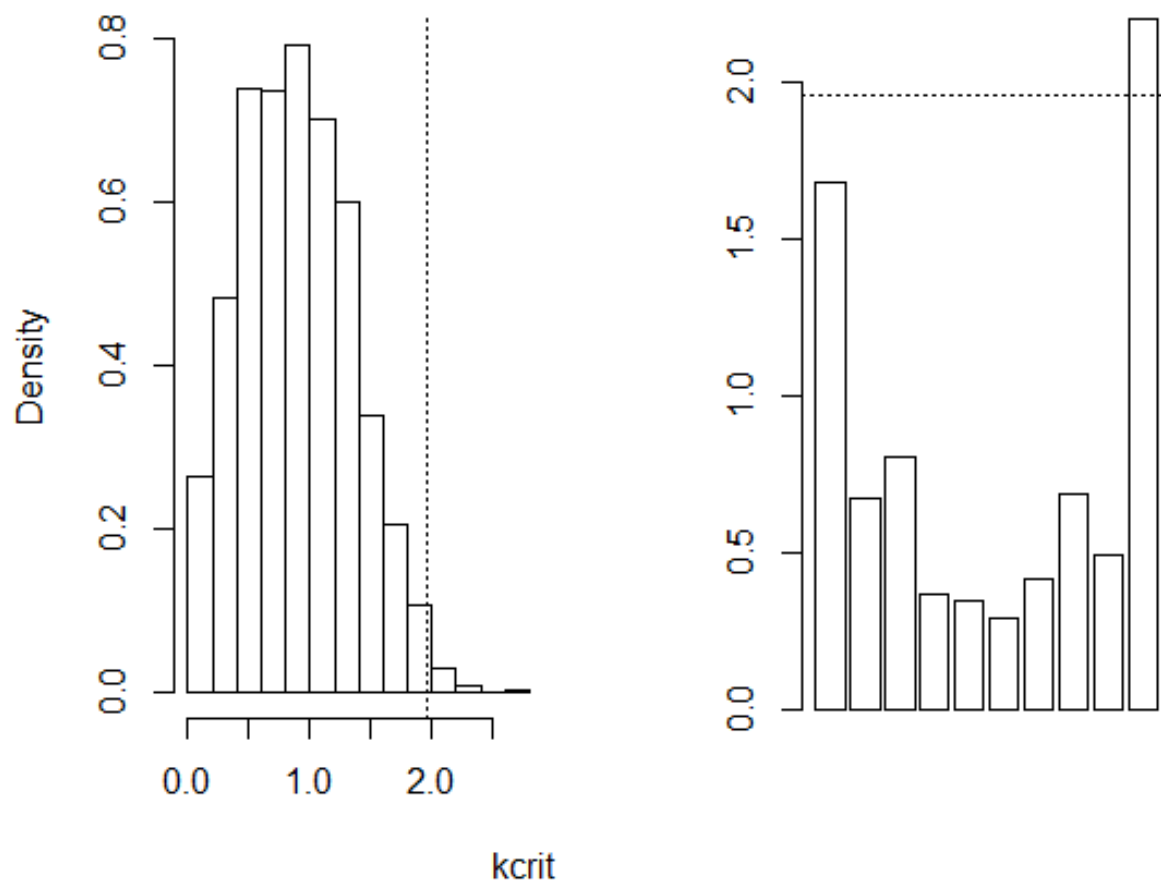
$L$	$n$	Estadístico $k$ Bootstrap ( $\Delta_\theta$ )							Estadístico $k$ univariante ( $\Delta_\theta$ )						
		(0)	(0,5)	(1)	(1,5)	(2)	(2,5)	(3)	(0)	(0,5)	(1)	(1,5)	(2)	(2,5)	(3)
5	3	0.013	0.082	0.206	0.320	0.420	0.529	0.609	0.006	0.070	0.187	0.298	0.410	0.518	0.595
	6	0.016	0.183	0.483	0.683	0.791	0.902	0.932	0.010	0.144	0.443	0.650	0.778	0.898	0.928
10	3	0.021	0.097	0.263	0.408	0.572	0.651	0.705	0.013	0.083	0.238	0.387	0.537	0.633	0.684
	6	0.020	0.234	0.564	0.756	0.880	0.923	0.954	0.009	0.190	0.513	0.717	0.861	0.913	0.946

En la tabla (4) se han registrado las proporciones de rechazo de la hipótesis nula cuando los datos siguen la distribución normal comparando el método bootstrap con el método tradicional, esta información se encuentra resumida en la figura (11), En la misma se puede comparar los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda), donde puede verse que conforme aumenta el tamaño  $n$  de las muestras (línea segmentada cuando  $n = 3$  y línea continua cuando  $n = 6$ ) aumenta la potencia de la prueba. También puede verse que el método bootstrap obtiene mejores resultados que el método tradicional.

La distribución bootstrap del estadístico para una de las corridas del algoritmo puede verse en la figura (12), puede verse como el último laboratorio ha sido detectado como atípico.



**Figura 11:** Curvas de potencia del estadístico  $k$  para la distribución normal.



**Figura 12:** La distribución bootstrap y el valor  $k$  crítico para la distribución normal.

### 5.3.2 Incumplimiento de la hipótesis nula de repetibilidad para la distribución de Laplace

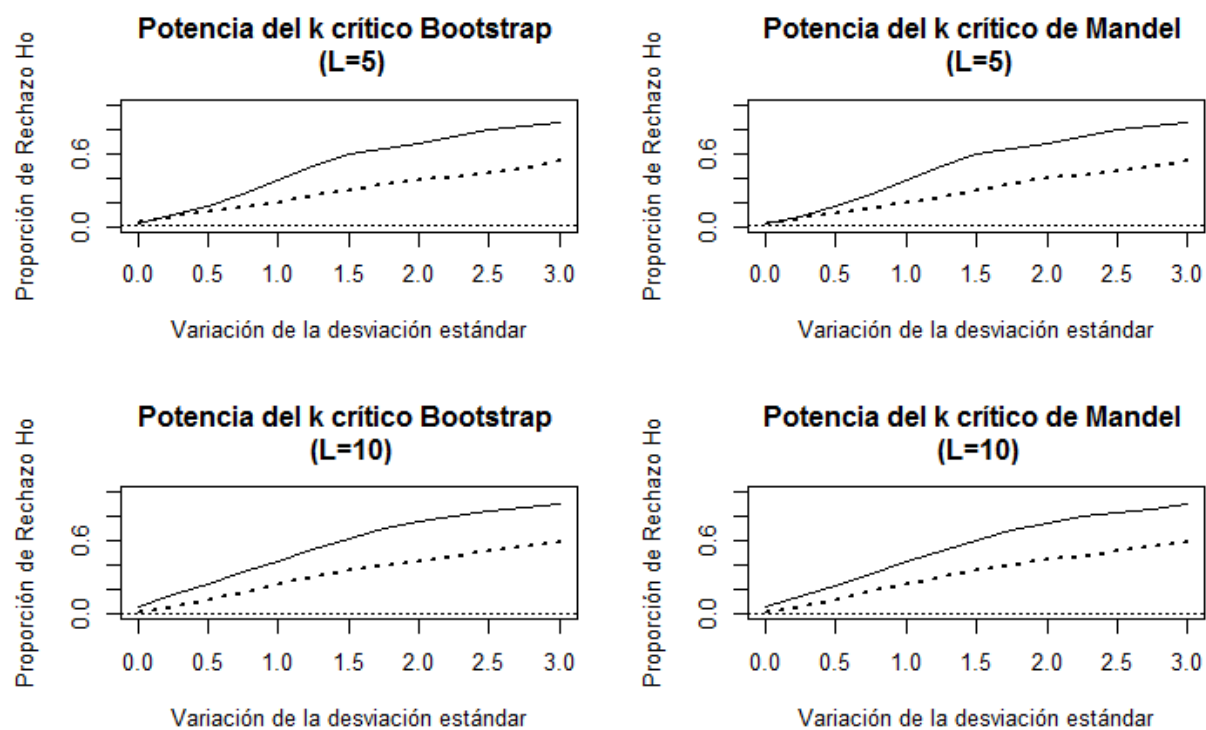
**Tabla 5:**

*Potencia del estadístico  $k$  para la distribución de Laplace*

$L$	$n$	Estadístico $k$ Bootstrap ( $\Delta_\theta$ )							Estadístico $k$ univariante ( $\Delta_\theta$ )						
		(0)	(0, 5)	(1)	(1, 5)	(2)	(2, 5)	(3)	(0)	(0, 5)	(1)	(1, 5)	(2)	(2, 5)	(3)
5	3	0.039	0.124	0.199	0.303	0.383	0.436	0.544	0.034	0.176	0.387	0.590	0.686	0.793	0.852
	6	0.034	0.176	0.387	0.590	0.686	0.793	0.852	0.028	0.165	0.390	0.592	0.688	0.795	0.856
10	3	0.023	0.121	0.250	0.361	0.435	0.525	0.596	0.021	0.122	0.250	0.364	0.438	0.525	0.593
	6	0.064	0.249	0.436	0.615	0.773	0.833	0.901	0.059	0.238	0.427	0.605	0.762	0.829	0.898

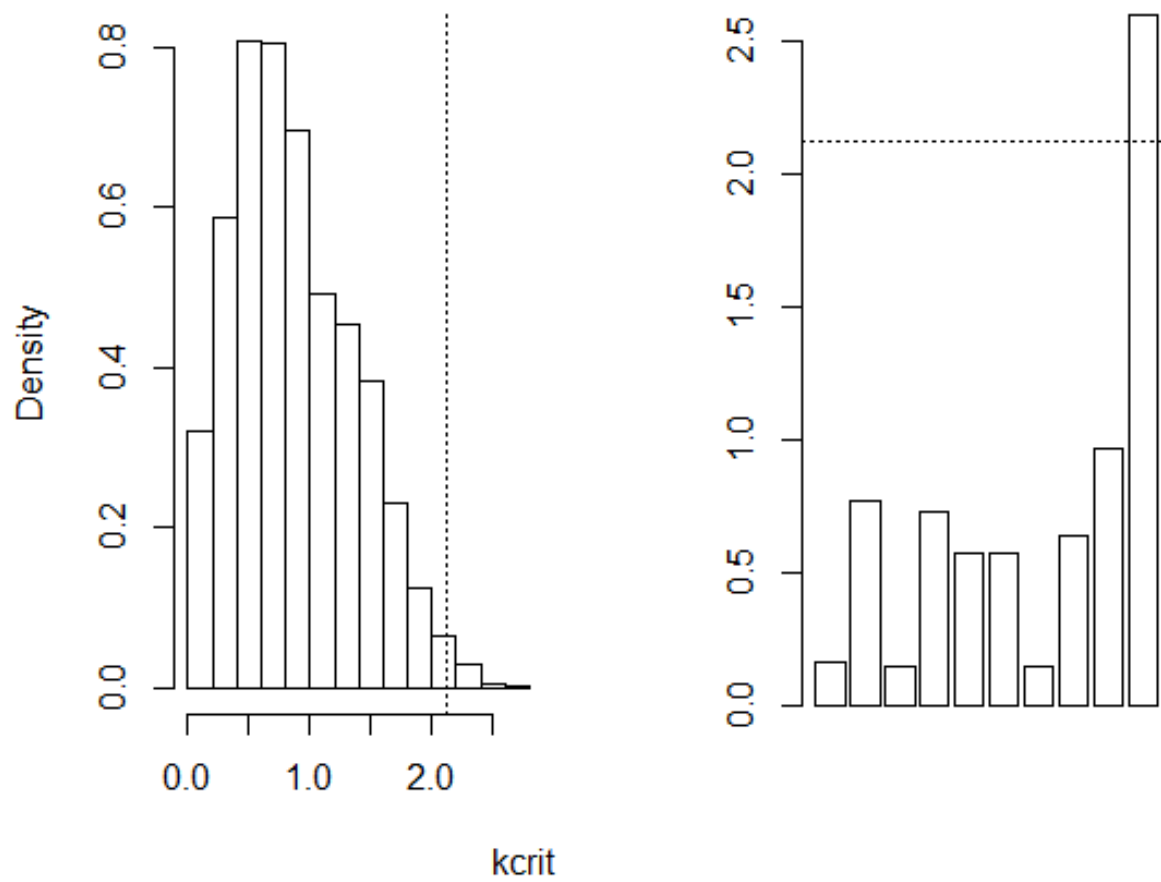
En la tabla (5) se han registrado las proporciones de rechazo de la hipótesis nula cuando los datos provienen de una distribución de Laplace comparando el método bootstrap con el método tradicional, esta información se encuentra resumida en la figura (13). En la misma se puede comparar los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda). Puede verse que conforme aumenta el tamaño  $n$  de las muestras (línea segmentada cuando  $n = 3$  y línea continua cuando  $n = 6$ ), aumenta la proporción de rechazo de la hipótesis nula. En general, puede verse que los resultados obtenidos por el método bootstrap y el método tradicional son similares.

La distribución bootstrap del estadístico  $k$  para una de las corridas del algoritmo puede verse en la figura (14, puede notarse como el último laboratorio ha sido detectado como atípico.



**Figura 13:** Curvas de potencia del estadístico  $k$  para la distribución de Laplace.





**Figura 14:** Distribución bootstrap y el valor  $k$  crítico para la distribución de Laplace.

### 5.3.3 Incumplimiento de la hipótesis nula de repetibilidad para la distribución sesgada

En la siguiente tabla se muestran los resultados del estudio de simulación aplicado a datos que siguen una distribución sesgada.

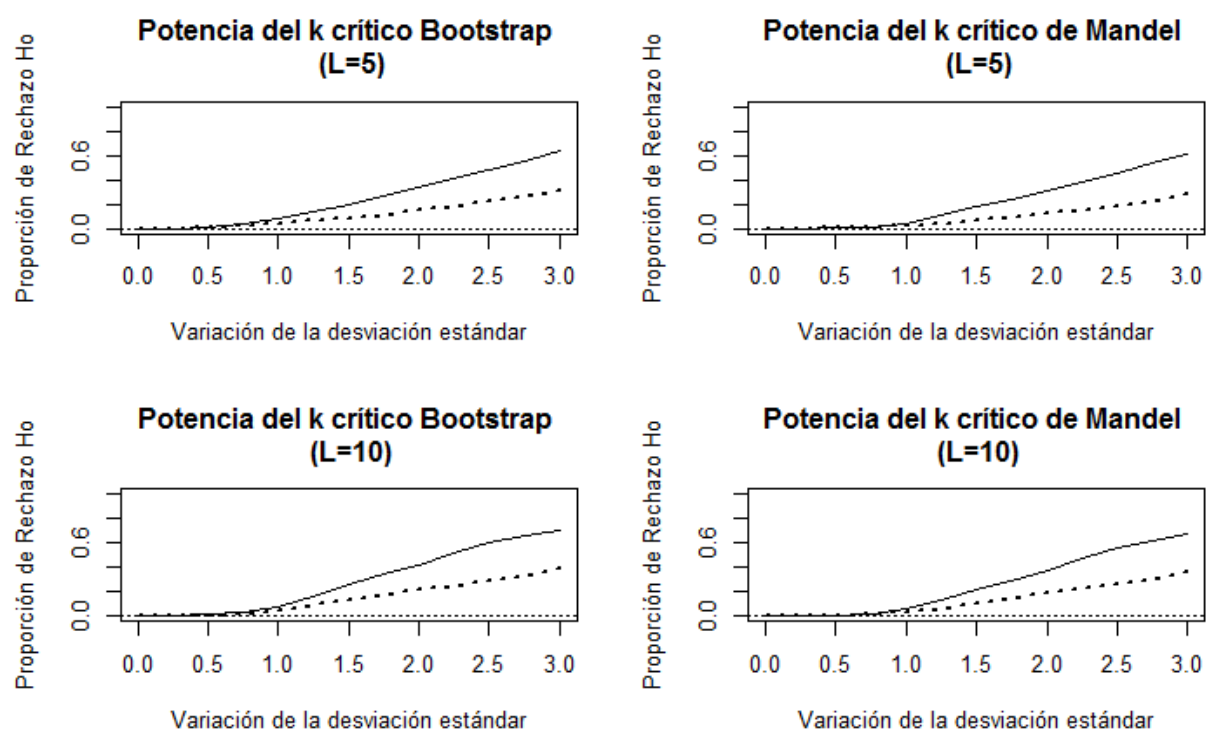
**Tabla 6:**

*Potencia del estadístico  $k$  para la distribución sesgada*

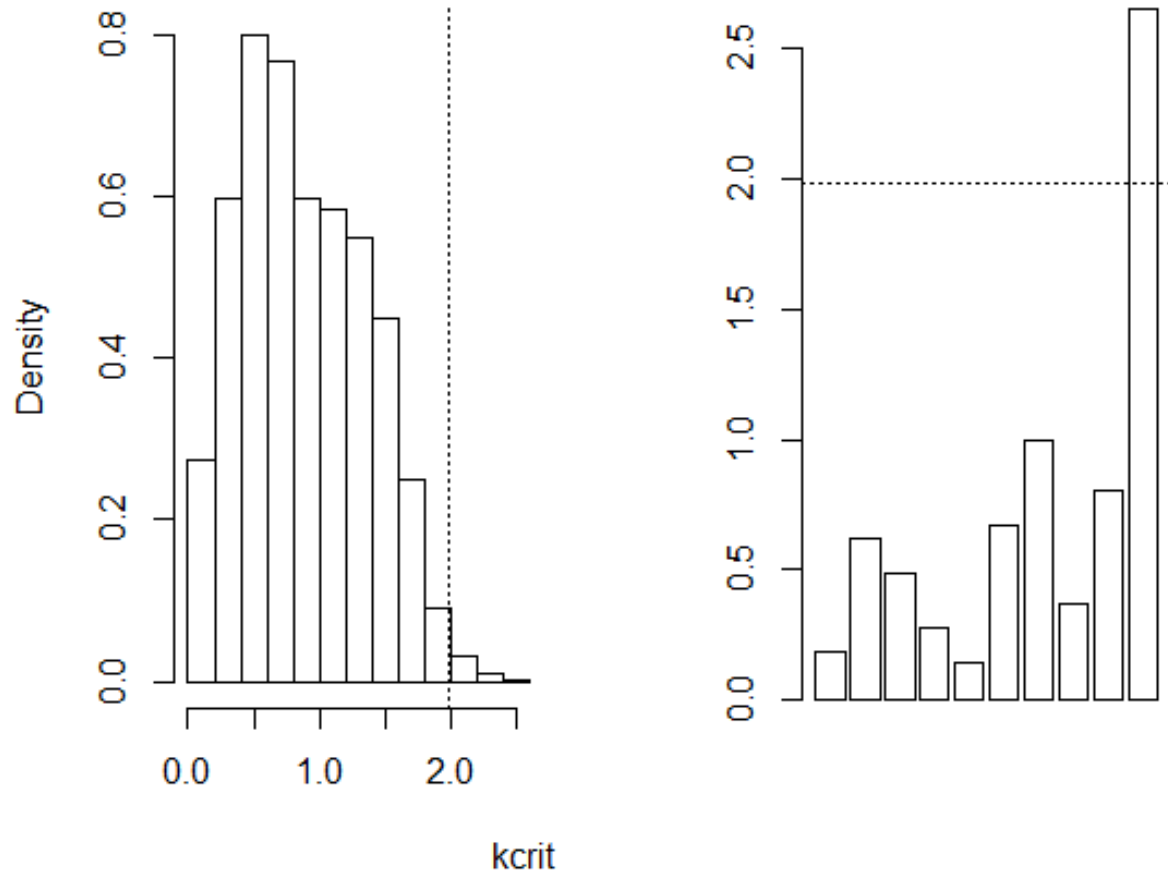
$L$	$n$	Estadístico $k$ Bootstrap ( $\Delta_\theta$ )							Estadístico $k$ univariante ( $\Delta_\theta$ )						
		(0)	(0,5)	(1)	(1,5)	(2)	(2,5)	(3)	(0)	(0.5)	(1)	(1.5)	(2)	(2.5)	(3)
5	3	0.005	0.014	0.046	0.088	0.150	0.221	0.317	0.001	0.008	0.033	0.065	0.124	0.184	0.288
	6	0.003	0.012	0.061	0.216	0.334	0.492	0.631	0.001	0.007	0.047	0.177	0.308	0.458	0.612
10	3	0.002	0.012	0.051	0.135	0.215	0.288	0.384	0.002	0.008	0.033	0.103	0.192	0.266	0.356
	6	0.002	0.018	0.083	0.258	0.425	0.601	0.705	0.001	0.013	0.060	0.214	0.382	0.560	0.670

En la tabla (6) se han registrado las proporciones de rechazo de la hipótesis nula para la distribución sesgada, que fue introducida en el capítulo anterior, comparando el método bootstrap con el método tradicional. Estos datos se encuentran resumidos en la figura (15). En la misma se puede comparar los resultados del estadístico  $h$  de Mandel obtenidos por el método tradicional (columna derecha) y por el método bootstrap propuesto (columna izquierda), se observa que para tamaños pequeños de muestras (línea segmentada cuando  $n = 3$  y línea continua cuando  $n = 6$ ), el método bootstrap detecta mejor los datos atípicos que el método tradicional.

La distribución bootstrap del estadístico para una de las corridas del algoritmo se encuentra en la figura (16), se observa que el último laboratorio ha sido detectado como atípico.



**Figura 15:** Curvas de potencia del estadístico  $k$  para la distribución sesgada.



**Figura 16:** La distribución bootstrap y el valor  $k$  crítico para la distribución sesgada.

## **6 Aplicación de la metodología propuesta**

### **6.1 Estudio Interlaboratorio para evaluar la precisión del método empleado para realizar exámenes sanguíneos de Biometría Hemática**

En esta sección aplicaremos la metodología propuesta a un conjunto de datos obtenidos de un Estudio Interlaboratorio que fue llevado a cabo como parte de esta investigación, luego compararemos la metodología propuesta con la metodología tradicional. El Estudio Interlaboratorio tuvo como objetivo determinar la precisión del método de medición utilizado para realizar exámenes sanguíneos mediante técnica computarizada, siguiendo las políticas y procedimientos establecidos en las normas ASTM E-691 e ISO 5725.

Existen diferentes tipos de instrumentos computarizados capaces de analizar los valores requeridos en una Biometría Hemática. Debido a la heterogeneidad de instrumentos utilizados, cada laboratorio mide diferentes aspectos en este tipo de examen, por ejemplo, encontramos que hubo laboratorios que midieron 21 propiedades sanguíneas, mientras otros midieron una cantidad menor de propiedades.

Para este estudio hemos seleccionado las características comunes que fueron consideradas por los nueve laboratorios que participaron del Estudio Interlaboratorio, con lo cual se conformaron 11 niveles de medición, o materiales, por valorar dentro de la misma muestra.

Los 11 niveles evaluados durante el Estudio Interlaboratorio fueron:

- (A) nivel de Hemoglobina.
- (B) Nivel de Hematocrito.
- (C) Cantidad de Glóbulos Rojos.
- (D) Cantidad de Glóbulos Blancos.

- (E) Porcentaje de Neutrófilos.
- (F) Porcentaje de Linfocitos.
- (G) Porcentaje de Células Intermedias.
- (H) Hemoglobina globular media.
- (I) Concentración media de hemoglobina modular.
- (J) Volumen globular medio.
- (K) Cantidad de Plaquetas.

Nuestro objetivo es sentar las bases para crear una cultura de control de la variación dentro de los laboratorios del país, y para ellos hemos considerado como población de estudio un grupo de laboratorios clínicos de la ciudad de Ibarra.

En la actualidad existen dos paquetes disponibles en el repositorio del software **R** que se encargan de realizar los cálculos de un Estudio Interlaboratorio mediante el procedimiento detallado en las normas ISO 5725 y ASTM E-691: el paquete *ILS* versión 0.2 (Flores et al., 2018) y el paquete *metRology* (Ellison, 2018) versión 0.9-28-1, disponibles en <https://cran.r-project.org>.

A continuación presentamos un resumen de los resultados del estudio interlaboratorio llevado a cabo, los cuáles fueron procesados en el software **R** a través del paquete *metRology*.

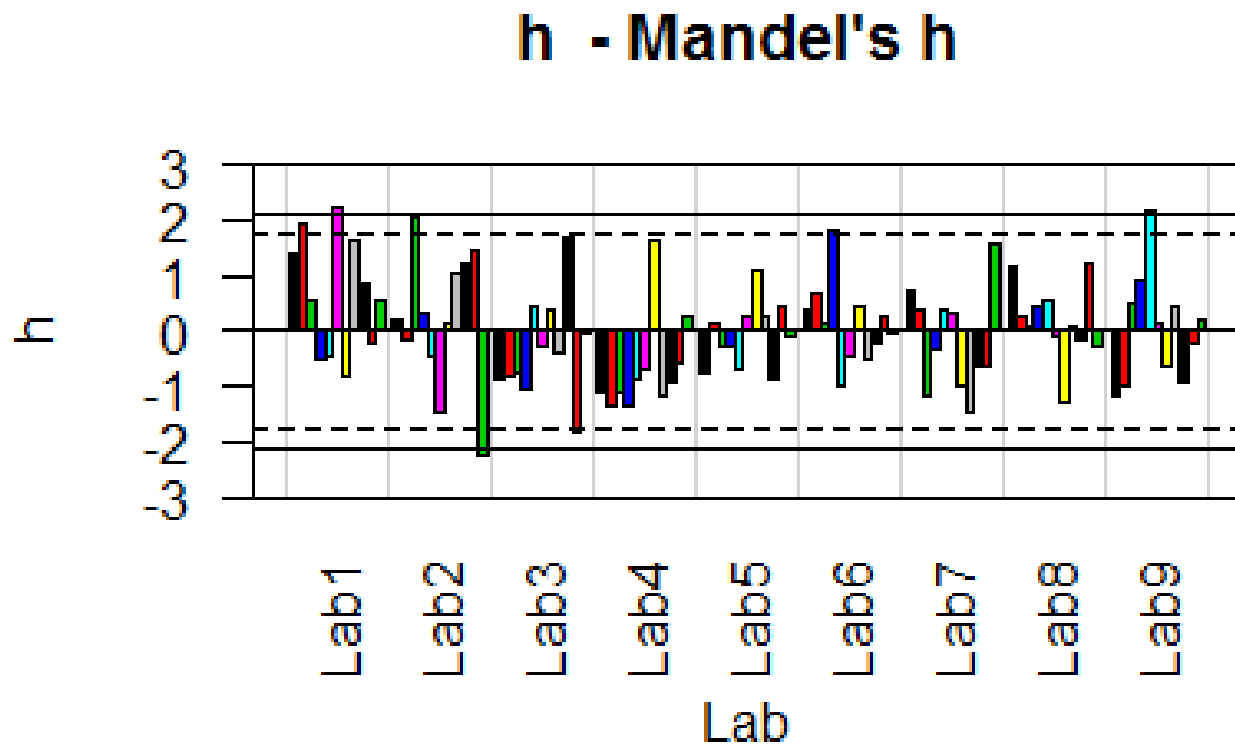
### **6.1.1 Estudio Interlaboratorio realizado por el método tradicional**

En la tabla (7) se muestran los estadísticos  $h$  de Mandel que fueron calculados mediante el paquete *metRology* para cada uno de los laboratorios y niveles.

**Tabla 7:***Estadísticos h de Mandel*

Laboratorio	Material										
	A	B	C	D	E	F	G	H	I	J	K
Lab1	1.38	1.94	0.57	-0.50	-0.48	2.23	-0.80	1.64	0.85	-0.20	0.54
Lab2	0.23	-0.18	2.04	0.31	-0.46	-1.44	0.16	1.04	1.20	1.47	-2.22
Lab3	-0.88	-0.82	-0.77	-1.07	0.41	-0.29	0.38	-0.37	1.70	-1.81	-0.04
Lab4	-1.10	-1.37	-1.10	-1.35	-0.90	-0.68	1.63	-1.15	-0.91	-0.58	0.27
Lab5	-0.76	0.16	-0.30	-0.29	-0.67	0.24	1.11	0.28	-0.88	0.42	-0.09
Lab6	0.38	0.67	0.16	1.83	-1.01	-0.44	0.47	-0.54	-0.23	0.26	-0.03
Lab7	0.76	0.37	-1.16	-0.36	0.40	0.34	-0.98	-1.48	-0.65	-0.61	1.60
Lab8	1.17	0.24	0.06	0.47	0.54	-0.09	-1.31	0.11	-0.15	1.24	-0.27
Lab9	-1.18	-0.99	0.49	0.95	2.16	0.13	-0.65	0.46	-0.93	-0.19	0.23

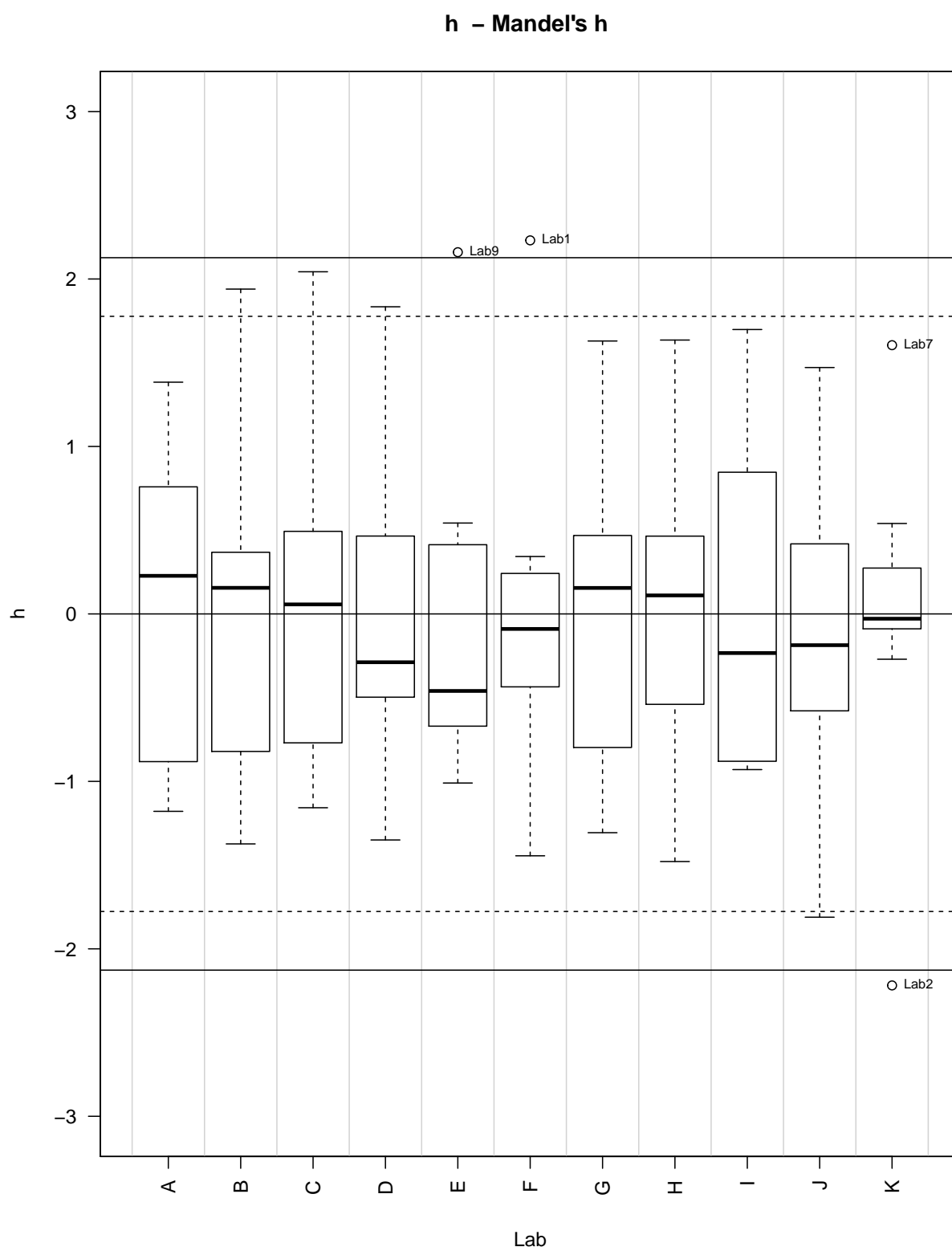
En la figura (17) vemos que los laboratorios 1, 2 y 9 presentan valores atípicos en los niveles  $F, K$  y  $E$  respectivamente. Mientras que el laboratorio cuatro muestra un sesgo negativo en los resultados, lo que debería ser investigado y reportado al laboratorio respectivo.



**Figura 17:** Diagrama de barras de los estadísticos  $h$  de Mandel.

En la figura (18) presentamos un diagrama de caja de los estadísticos  $h$  de Mandel por cada nivel de análisis, podemos ver los datos atípicos antes mencionados por encima del valor crítico del estadístico  $h$  para un  $\alpha = 0.01$ .





**Figura 18:** Diagramas de caja de los estadísticos  $h$  de Mandel.

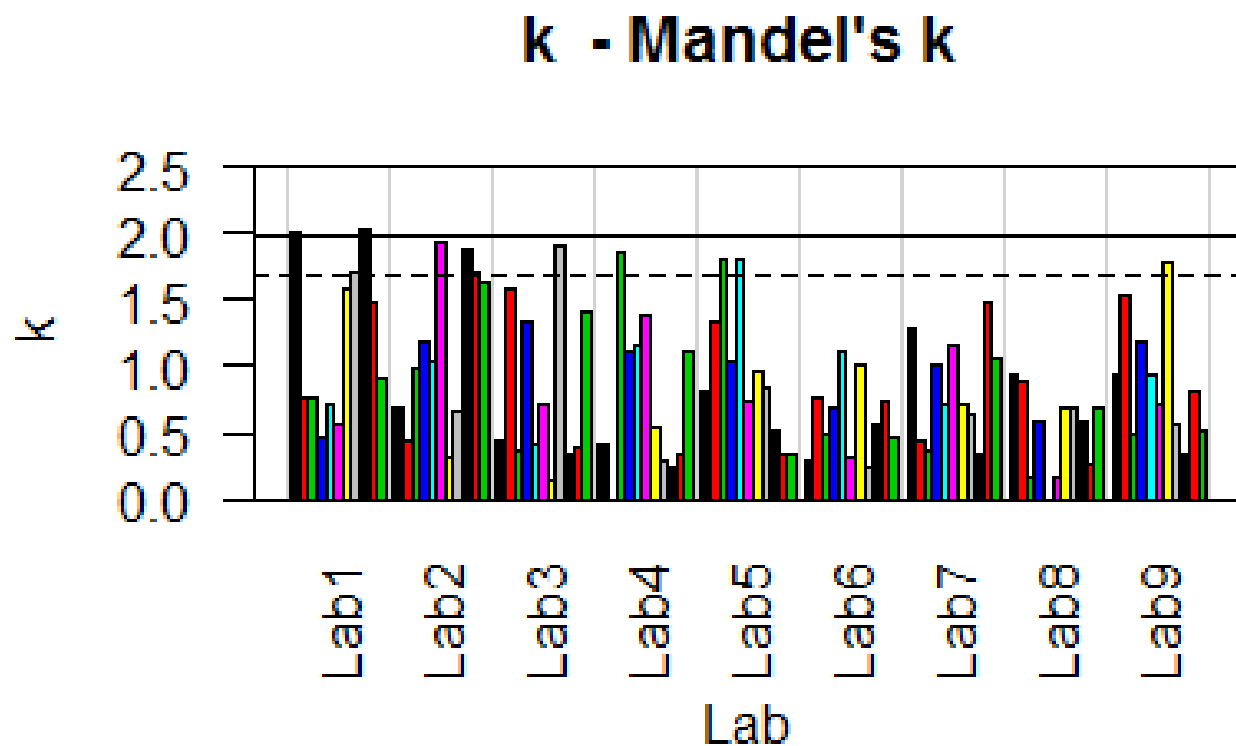
En la tabla (8) se presentan los estadísticos  $k$  de Mandel calculados mediante el paquete *ILS*.

**Tabla 8:**

*Estadísticos  $k$  de Mandel.*

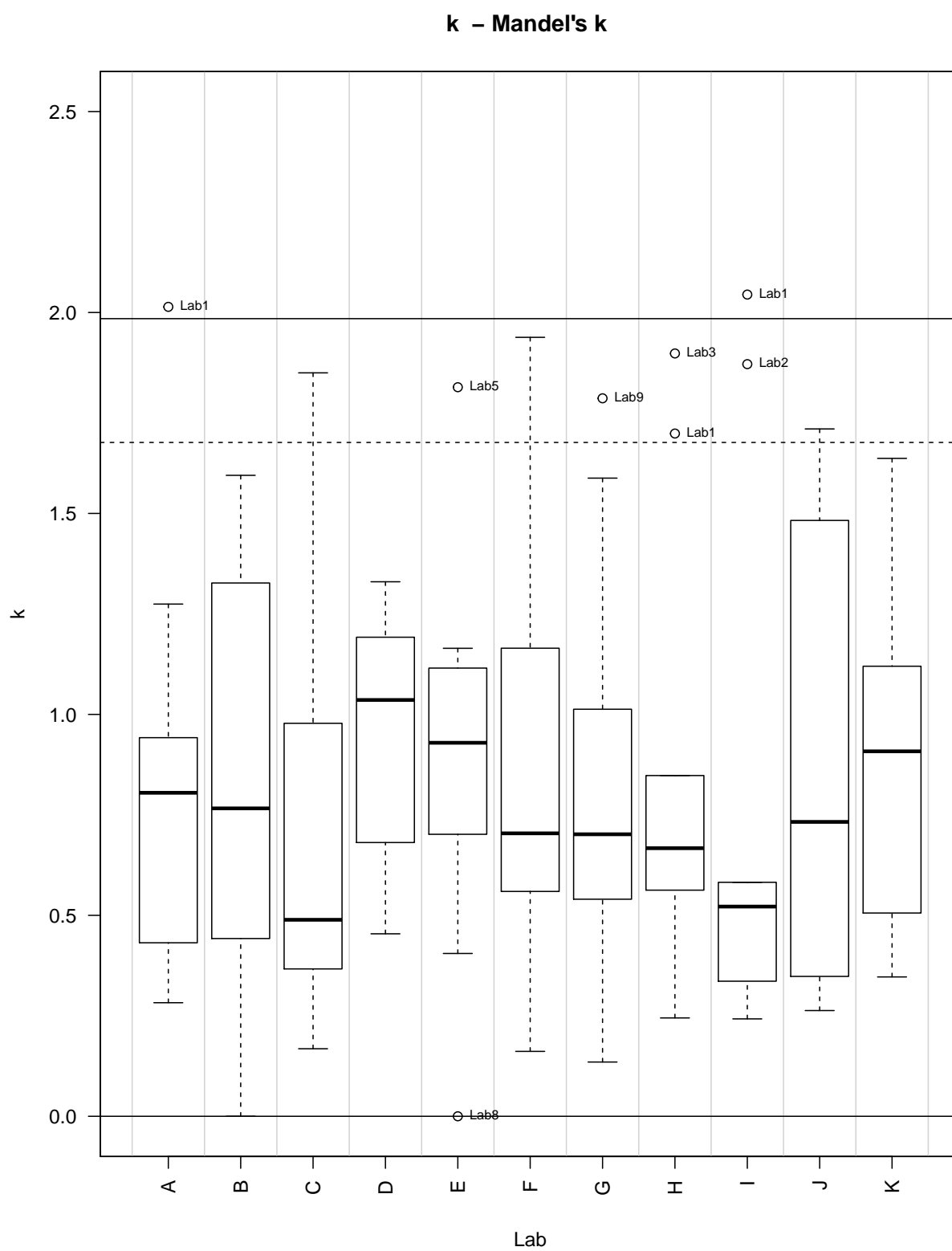
Material											
Laboratorio	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
Lab1	2.01	0.77	0.76	0.45	0.70	0.56	1.59	1.70	2.04	1.49	0.91
Lab2	0.68	0.44	0.98	1.19	1.05	1.94	0.31	0.67	1.87	1.71	1.64
Lab3	0.43	1.59	0.37	1.33	0.41	0.70	0.14	1.90	0.34	0.39	1.42
Lab4	0.41	0.00	1.85	1.12	1.16	1.38	0.54	0.28	0.24	0.35	1.20
Lab5	0.80	1.32	1.81	1.04	1.81	0.74	0.95	0.85	0.52	0.35	0.35
Lab6	0.28	0.77	0.49	0.68	1.12	0.32	1.01	0.24	0.56	0.73	0.47
Lab7	1.27	0.44	0.37	1.02	0.70	1.16	0.70	0.65	0.34	1.48	1.07
Lab8	0.94	0.88	0.17	0.60	0.00	0.16	0.68	0.69	0.58	0.26	0.69
Lab9	0.94	1.53	0.49	1.19	0.93	0.70	1.79	0.56	0.34	0.80	0.51

La figura (19) corresponde a la representación de los estadísticos  $k$  de Mandel, podemos ver que el laboratorio 1 presenta datos atípicos en los niveles *A* e *I*, por lo que este laboratorio debería ser investigado.



**Figura 19:** Diagrama de barras de los Estadísticos  $k$  de Mandel.

En la figura (20) presentamos un diagrama de caja de los estadísticos  $k$  de Mandel por cada nivel de análisis, podemos ver los datos atípicos por encima del valor crítico del estadístico  $k$  para un  $\alpha = 0.01$ .



**Figura 20:** Diagramas de caja de los estadísticos  $k$  de Mandel.

Posteriormente es necesario aplicar las pruebas de Grubbs y de Cochran para verificar si los valores atípicos detectados deben ser apartados del resto de datos, sin embargo, no hemos realizado este cálculo ya que el propósito de la presente investigación fue centrarse en la detección de datos atípicos.

Finalmente, los valores de la desviación estándar de repetibilidad  $S_r$ , de la desviación estándar de reproducibilidad  $S_R$ , y de la media general para cada nivel  $m$ , se presentan en la tabla (9). Estos valores fueron calculados con el paquete *ILS*. Podemos ver que no existe una aparente dependencia funcional entre  $m$  y  $S_r$ , o entre  $m$  y  $S_R$ .

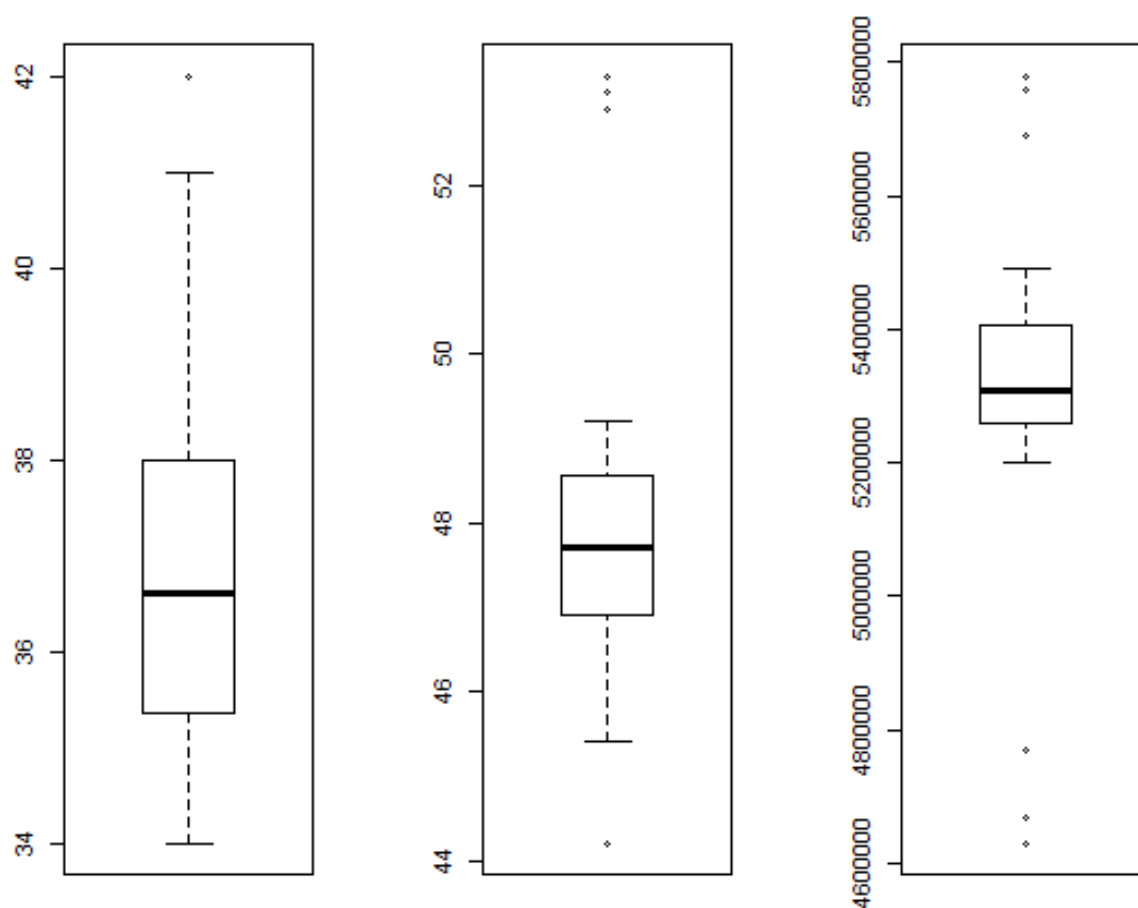
**Tabla 9:**

*Desviaciones estándar de repetibilidad y reproducibilidad*

Nivel	$m$	$S$	$S_r$	$S_L$	$S_R$
A	8.18	0.327	0.613	2.13	2.19
B	16.08	0.0702	0.131	0.78	0.79
C	30.99	0.295	0.472	2.32	2.35
D	33.64	0.103	0.339	0.94	0.98
E	36.88	0.318	0.621	2.06	2.12
F	47.94	0.198	0.357	2.31	2.33
G	55.00	0.469	0.855	2.55	2.64
H	90.85	0.362	0.624	4.05	4.08
I	6469.63	0.0119	0.0172	469.30	489.81
J	295888.9	0.00251	0.00439	29754.55	29969.5
K	5301222	0.000195	0.00044	275502.8	277840.5

### 6.1.2 Estudio Interlaboratorio realizado por el método *bootstrap*

Como primer paso de la metodología propuesta se deben detectar y luego recortar los datos atípicos con la ayuda de un diagrama de caja. Como resultado de esta operación en el material  $E$  se encontró un dato atípico, en el material  $F$  se encontraron cuatro datos atípicos, y en los resultados del material  $K$  se encontraron seis datos atípicos, como puede verse en la figura (21).



**Figura 21:** Diagramas de caja de los datos provenientes de los laboratorios  $E$ ,  $F$  y  $K$ .

A continuación presentamos los valores críticos del estadístico  $h$  que fueron determinados por el método bootstrap cuando hemos empleado los datos del Estudio Interlaboratorio, para ello se

realizaron  $B = 1000$  remuestras de las mediciones hechas para cada material, para cada remuestra se calculó el valor del estadístico  $h$  con lo cual se generó la distribución bootstrap del estadístico.

Podemos notar en la tabla (10) que los valores críticos correspondientes al estadístico  $h$  no son simétricos lo que, a diferencia del método tradicional, es característico de la distribución bootstrap.

**Tabla 10:**

*Valores críticos de los estadísticos  $h$  bootstrap.*

Nivel	$h_{1-\frac{\alpha}{2}}$	$h_{\frac{\alpha}{2}}$
A	-2.02	2.15
B	-2.04	2.11
C	-1.96	2.20
D	-2.06	2.14
E	-1.97	2.24
F	-2.19	2.04
G	-2.06	2.14
H	-2.12	2.10
I	-1.94	2.16
J	-2.17	2.11
K	-2.02	2.15

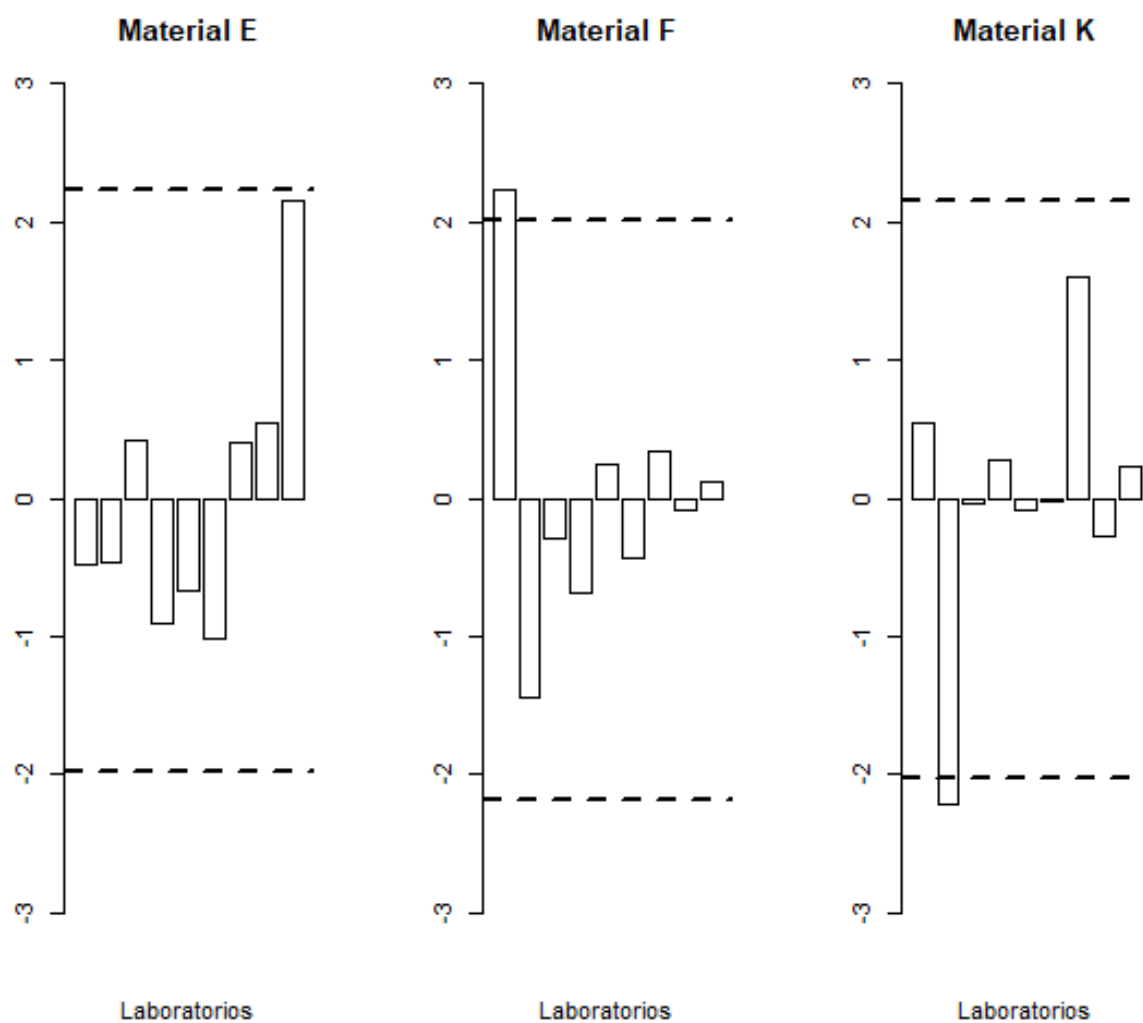
En la tabla (11) se han resaltado los valores  $h$  que están por encima de su respectivo valor crítico.

**Tabla 11:***Estadísticos  $h$  bootstrap*

Material											
Laboratorio	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
Lab1	1.38	1.94	0.57	-0.50	-0.48	<b>2.23</b>	-0.80	1.64	0.85	-0.20	0.54
Lab2	0.23	-0.18	2.04	0.31	-0.46	-1.44	0.16	1.04	1.20	1.47	<b>-2.22</b>
Lab3	-0.88	-0.82	-0.77	-1.07	0.41	-0.29	0.38	-0.37	1.70	-1.81	-0.04
Lab4	-1.10	-1.37	-1.10	-1.35	-0.90	-0.68	1.63	-1.15	-0.91	-0.58	0.27
Lab5	-0.76	0.16	-0.30	-0.29	-0.67	0.24	1.11	0.28	-0.88	0.42	-0.09
Lab6	0.38	0.67	0.16	1.83	-1.01	-0.44	0.47	-0.54	-0.23	0.26	-0.03
Lab7	0.76	0.37	-1.16	-0.36	0.40	0.34	-0.98	-1.48	-0.65	-0.61	1.60
Lab8	1.17	0.24	0.06	0.47	0.54	-0.09	-1.31	0.11	-0.15	1.24	-0.27
Lab9	-1.18	-0.99	0.49	0.95	<b>2.16</b>	0.13	-0.65	0.46	-0.93	-0.19	0.23

Como se puede ver en la figura (22), el noveno laboratorio no fue detectado como atípico para el material *E* como en el método tradicional. Por otra parte, el primer laboratorio para el nivel *F*, y el segundo laboratorio para el nivel *K*, fueron marcados como laboratorios con datos atípicos.





**Figura 22:** Valores  $h$  críticos bootstrap y los laboratorios 1, y 2 detectados como atípicos.

En la tabla (12) presentamos los valores críticos del estadístico  $k$  que fueron determinados por el método bootstrap cuando hemos empleado los datos del caso práctico. Podemos ver que los valores críticos para los distintos niveles no son iguales, a diferencia del valor crítico  $k$  del método tradicional, cuyo valor permanece constante dado número de laboratorios y el número de réplicas.

**Tabla 12:**

*Valores críticos del estadístico  $k$  bootstrap*

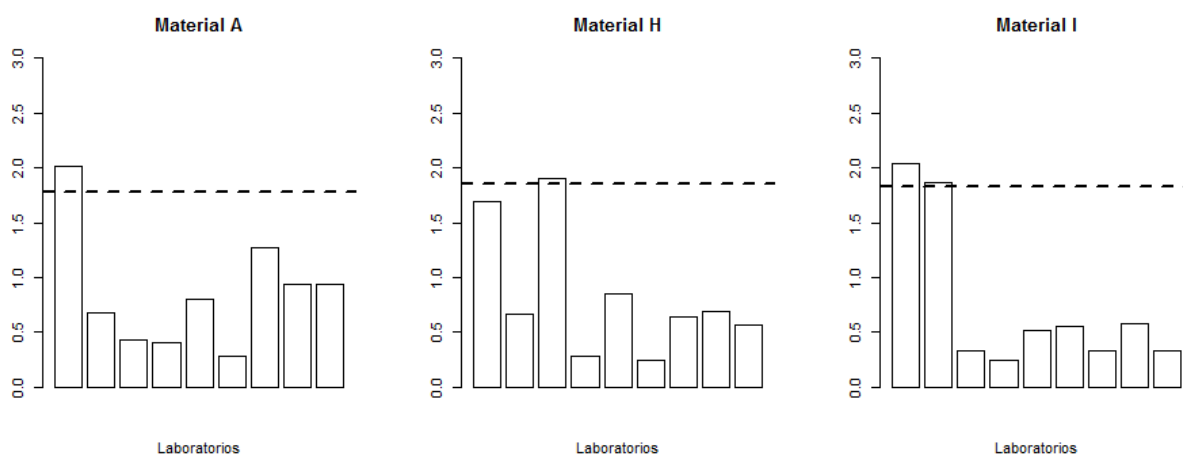
Nivel	$k_\alpha$
A	1.78
B	1.94
C	1.97
D	1.90
E	2.01
F	1.96
G	1.80
H	1.86
I	1.83
J	1.92
K	1.87

En la tabla (13) podemos ver resaltadas las puntuaciones  $k$  que han sido detectadas como inconsistentes.

**Tabla 13:***Estadísticos  $k$  bootstrap*

Laboratorio	Material										
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
Lab1	<b>2.01</b>	0.77	0.76	0.45	0.70	0.56	1.59	1.70	<b>2.04</b>	1.49	0.91
Lab2	0.68	0.44	0.98	1.19	1.05	1.94	0.31	0.67	<b>1.87</b>	1.71	1.64
Lab3	0.43	1.59	0.37	1.33	0.41	0.70	0.14	<b>1.90</b>	0.34	0.39	1.42
Lab4	0.41	0.00	1.85	1.12	1.16	1.38	0.54	0.28	0.24	0.35	1.20
Lab5	0.80	1.32	1.81	1.04	1.81	0.74	0.95	0.85	0.52	0.35	0.35
Lab6	0.28	0.77	0.49	0.68	1.12	0.32	1.01	0.24	0.56	0.73	0.47
Lab7	1.27	0.44	0.37	1.02	0.70	1.16	0.70	0.65	0.34	1.48	1.07
Lab8	0.94	0.88	0.17	0.60	0.00	0.16	0.68	0.69	0.58	0.26	0.69
Lab9	0.94	1.53	0.49	1.19	0.93	0.70	1.79	0.56	0.34	0.80	0.51

Podemos ver que se detectarán como atípicos el primer laboratorio para el nivel  $K$ , el tercer laboratorio para el nivel  $H$ , y el primer y segundo laboratorio para el nivel  $I$ . Estos laboratorios deben ser investigados para tratar de explicar las causas de la variabilidad y detectar posibles errores en la aplicación de un método de medición.



**Figura 23:** Valores  $k$  críticos bootstrap y los laboratorios 1 y 2 detectados como atípicos.

Hay que anotar que este trabajo se limitó a señalar los laboratorios que presentan valores atípicos, y no se procede a completar todo el proceso de análisis que requiere un Estudio Interlaboratorio.

## 7 Conclusiones y recomendaciones

En el presente trabajo se ha propuesto una nueva metodología para detectar resultados atípicos en Estudios Interlaboratorio mediante la aplicación del método bootstrap. Este nuevo enfoque fue capaz de detectar laboratorios atípicos sin recurrir a los supuestos de normalidad sobre las variables en estudio, lo que permitió desarrollar una metodología alternativa para detectar resultados atípicos en Estudios Interlaboratorio.

Las distribuciones de probabilidad de los estadísticos  $h$  y  $k$  se estimaron mediante el procedimiento bootstrap propuesto, con el cual fue posible obtener los valores críticos de los estadísticos  $h$  y  $k$  como los cuantiles de las correspondientes distribuciones bootstrap de los estadísticos para un nivel de significación de  $\alpha = 0.01$ .

Para obtener la distribución bootstrap de los estadísticos  $h$  y  $k$  fue necesario realizar remuestreo a partir de la hipótesis nula de que todos los laboratorios provienen de la misma población, motivo por el que se procedió a separar los resultados que fueron detectados como atípicos mediante un diagrama de caja aplicado a los datos. Este procedimiento demostró ser útil ya que fue posible detectar resultados atípicos en una variedad de escenarios, por ejemplo, en los escenarios de simulación que se comprobó que la potencia de la prueba obtenida por el método bootstrap fue igual o superior a la potencia de la prueba del método tradicional.

Durante el estudio de simulación, se comprobó que los laboratorios que fueron simulados como atípicos fueron detectados exitosamente mediante la metodología propuesta. También se verificó que los cambios en la media de los laboratorios fueron detectados mediante el estadístico  $h$ , mientras que los cambios en la varianza de los laboratorios fueron detectados por el estadístico  $k$ .

En cuanto a la potencia del método, se ha verificado que cuando el tamaño de las muestras no es muy alto, y la distribución considerada es asimétrica, el método bootstrap aplicado al estadístico  $k$  ofrece mejores resultados que el método tradicional. La metodología propuesta hace posible que

se puedan detectar resultados atípicos evitando plantear los supuestos sobre las variables aleatorias en estudio.

## BIBLIOGRAFÍA

- Amat, R. (2016). Resampling: Test de permutación, Simulación de Monte Carlo y Bootstrapping.
- ASTM (2005). *Standard practice for conducting an interlaboratory study to determine the precision of a test method*. American Society for Testing and Materials. ASTM E-691 Standard.
- Barrentine, L. (2003). *Concepts for R&R Studies*. ASQ Quality Press.
- Cochran, W. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen*, 11:47–52.
- Ellison, S. (2018). *The R Package metRology: Support for Metrological Applications*.
- Figueireido, F. and Gomes, I. M. (2013). The Skew-Normal Distribution in SPC. *REVSTAT - Statistical Journal*, 11(1):83–104.
- Flores, M. (2018). *Nuevas Aportaciones del Análisis de Datos Funcionales en el Control Estadístico de Procesos*. PhD thesis, Departamento de Matemáticas. Universidade da Coruña.
- Flores, M., Fernandez, R., Naya, S., and Tarrío-Saavedra, J. (2018). *The R Package ILS: Interlaboratory Study*.
- Grubbs, F. (1950). Sample criteria for testing outlying observations. *Ann. Math. Stat*, 21:27–58.
- Hesterberg, T. (2003). *Bootstrap Methods and Permutation Tests*. W.H. Freeman and Company.
- Highways, S. and Transpotation (1996). *Procedures Manual for Interlaboratory Testing Programs*. Saskatchewan Highways and Transpotation. STP 304-3 Standard.

- ISO (1994). *Accuracy - Trueness and Precision - of Measurement Methods and Results. Part 2: Methods for the Determination of Repeteability and Reproducibility*. International Organization for Standardization. ISO 5725 Standard.
- Lyndsey, J. and Swihart, B. (2019). *The R Package Manual rmutil: Utilities for Nonlinear Regression and Repeated Measurements Models*.
- Navidi, W. (2006). *Estadística para Ingenieros y Científicos*. McGraw-Hill.
- Pastoriza, I. (2013). Estudios Interlaboratorio. Notas entregadas en el Máster en Técnicas Estadísticas de UVIGO, USC y UDC.
- Solanas, A. y Sierras, V. (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de Psicología*, 55:143–154.
- Vander Heyden, Y. and Smeyers-Verbeke, J. (2007). Set-up and evaluation of interlaboratory studies. *Journal of Chromatography*, A(1158):158–167.
- Wilrich, P. (2013). Critical values of Mandel's h and k, the Grubbs and the Cochran test statistic. *AStA Adv Stat Anual*, 97:1–10.