METODOLOGÍA PARA LA EVALUACIÓN DE LOS INDICADORES EFICACIA Y PÉRDIDAS ASOCIADOS A LA RECOLECCIÓN DE CAFÉ

ANDRÉS ALBERTO GUTIÉRREZ MONTEALEGRE

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERIA INDUSTRIAL Y ESTADÍSTICA
SANTIAGO DE CALI
2006

METODOLOGÍA PARA LA EVALUACIÓN DE LOS INDICADORES EFICACIA Y PÉRDIDAS ASOCIADOS A LA RECOLECCIÓN DE CAFÉ

ANDRÉS ALBERTO GUTIÉRREZ MONTEALEGRE

Trabajo de grado para optar
por el título de Estadístico

Director
Jaime Enrique Pérez Restrepo
Ingeniero Civil
M.Sc. Estadística Matemática

Co-Director
Esther Cecilia Montoya Restrepo
Estadístico
M. Sc. Investigación de Operaciones

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERIA
ESCUELA DE INGENIERIA INDUSTRIAL Y ESTADISTICA
SANTIAGO DE CALI
2006

A Dios

A mis padres

A Aura,
donde quiera que estés

AGRADECIMIENTOS

El autor expresa sus agradecimientos a:

- Dios.

- Mi madre y a mi abuela Aura (q. e. p. d.), por ser los alicientes de mi trabajo.

- Jaime Pérez, por su ayuda y su profesionalismo como director del trabajo.

- Centro Nacional de Investigaciones del Café, CENICAFÉ.

- Profesores del programa de Estadística de la Universidad del Valle, en especial a Marisol Gordillo y Roberto Behar, y a Héctor Fabio Ramírez, Facultad de Economía.

- Miryam Cristina Duque, CIAT y Fernando Montealegre, Universidad de Toronto.

- Hernando Duque, Luis Eduardo Isaza, Carlos Oliveros, Claudia Rocío Gómez, Jaime Arcila, Hernando García, Luis Ignacio Estrada, Blanca Vargas, Flor Pulido, en CENICAFÉ; y Juan Carlos García, Deisy Cuartas, Alberto Echeverry, Luis Carlos Sánchez, en la Estación Central Naranjal.

- María Mercedes Cuartas y Eduardo Mejía, propietarios de la finca La Soledad, municipio de Palestina.

• Henry Cortés Lara, María Clarisa Osorio y María Eva Morales, por su hospitalidad durante mi estadía en el municipio de Chinchiná.

• Juliana Naranjo, Estiven Valencia, Carmenza Gutiérrez, Albeiro Salazar, Rocío Pineda, Valentín Erazo, William Ocampo, Patricia Lopera, Hernán Otálvaro, Lady Vargas, Alex Atehortua.

CONTENIDO

Pag.

LISTA DE TABLAS

Pág.

LISTA DE FIGURAS

Pág.

LISTA DE ANEXOS

Pág.

RESUMEN

El objetivo del trabajo es proponer un método de muestreo para evaluar el número de frutos en el árbol (eficacia), y el número de frutos en el suelo (pérdidas), después de una jornada de recolección de café. Teniendo en cuenta la dificultad de realizar un muestreo aleatorio simple en campo, y la necesidad de obtener resultados rápidamente, se proponen dos muestreos alternativos que no están basados en la aleatoriedad: un recorrido en recta, y un recorrido en diagonal. Dado que estos dos recorridos no están basados en aleatoriedad, se realizó la comparación de los valores de cada uno de ellos, con los valores obtenidos por muestreo aleatorio simple mediante una prueba de aleatorización (Fisher, 1935; Pitman, 1937). El trabajo de campo se realizó en dos fincas del departamento de Caldas, evaluando cuatro jornadas en cada una de ellas. El recorrido en recta tuvo valores similares a los del muestreo aleatorio simple en tres de las cuatro comparaciones realizadas; mientras que el recorrido en diagonal solo tuvo valores similares en una de las cuatro comparaciones.

Palabras clave: Recolección de café, Eficacia, Pérdidas, Muestreo, Pruebas de aleatorización, Reordenamientos aleatorios, Estadísticos de prueba equivalentes, Distribución de referencia, Valor *p*.

INTRODUCCIÓN

El presente trabajo de grado corresponde a la investigación BIO 0309 de la disciplina de Biometría del Centro Nacional de Investigaciones del Café, CENICAFE. Hace parte de las investigaciones realizadas en esta disciplina sobre recolección manual de café. Dado que este proceso constituye el 42 % de los costos total de producción en Colombia (FEDERACAFÉ, 2000), se hacen necesarios estudios que permitan establecer mejoras en el proceso.

Con la realización del trabajo, se propone una metodología de muestreo para la estimación de los promedios por árbol, tanto del número de frutos maduros sin recolectar, como del número de frutos dejados en el suelo, después de una jornada de recolección. Estas variables están asociadas a los indicadores Eficacia y Pérdidas, respectivamente.

Se estudiaron dos métodos de selección de árboles para la evaluación de los indicadores, teniendo en cuenta el carácter práctico de su aplicación en campo, respecto al tiempo de aplicación, y la importancia de tomar a tiempo las decisiones sobre el proceso.

El trabajo de campo se realizó durante la cosecha de mitaca de 2005 en dos localidades del departamento de Caldas.

Se realizó un paralelo entre el muestreo aleatorio simple, y dos recorridos para tomar una muestra de árboles. El objetivo fue proponer al caficultor un método

para evaluar la recolección, que pueda realizarse diariamente y con base en los resultados obtenidos, si es el caso, establecer nuevas estrategias con el propósito de mejorar el proceso en términos de operación y costos.

1. DEFINICIÓN DEL PROBLEMA

1.1 PLANTEAMIENTO

La rentabilidad de la producción de café en Colombia depende de factores como el precio externo, que afecta el interno, la tasa de cambio y los costos de producción. De los factores anteriores, el caficultor solamente puede actuar disminuyendo al máximo sus costos de producción, de los cuales, los costos de la recolección constituyen los de mayor porcentaje sobre el total, representando aproximadamente un 42% (Federacafé, 2000).

La recolección del café en Colombia se realiza en forma manual desprendiendo de manera selectiva los frutos maduros. El alto costo de la recolección es debido a varios factores, entre ellos, la alta selectividad y eficacia exigidas, la maduración no uniforme que exige varios pases al año, el costo del salario mínimo legal, la edad de la plantación, la cercanía a la cabecera municipal y el trato dado en la finca, como la alimentación, alojamiento, recibo del café, entre otros.

La recolección manual de café es una actividad compleja, basada en conocimientos empíricos derivados de la propia experiencia y en ocasiones transmitidos de generación en generación. En el campo hay dos protagonistas fundamentales, cada uno con una función específica: el recolector, encargado de desprender el café y llevarlo hasta el sitio de acopio, y el patrón de corte, encargado de conseguir el personal, asignar el trabajo y supervisar la actividad, entre otras.

Las recomendaciones que se realizan a los recolectores se pueden resumir en tres aspectos, principalmente:

- Reducir el número de frutos maduros dejados en los árboles después de un pase.

- Reducir el número de frutos dejados en el suelo.

- Reducir el porcentaje de frutos verdes en los frutos recolectados.

Respecto a cada uno de estos ítems, existen estándares con los cuales se deben comparar los valores obtenidos y así poder determinar si la recolección se realizó o no de una manera adecuada. Estos estándares han sido resultado de investigaciones realizadas en Cenicafé. Para el manejo integrado de la broca del café (MIB), se recomienda que tanto para el número de frutos maduros medido por el indicador Eficacia, como para el número de frutos en el suelo medido por el indicador Pérdidas, los promedio por árbol sean menores a cinco frutos (Bustillo, 2002). Adicionalmente, los frutos dejados de recolectar constituyen un ingreso dejado de percibir por el caficultor (Duque, 2002).

Con el fin de obtener una producción de café de alta calidad, el porcentaje de frutos verdes en la masa cosechada (indicador Calidad) debe tomar valores menores de 2,5% (Puerta, 2000).

Sin embargo, en investigaciones realizadas en Cenicafé (Díaz y Marín, 1999; Villegas, 2003), se ha encontrado que la recolección de café con frecuencia no se ajusta a estándares deseables para el manejo integrado de la broca del café (MIB),

es decir, los promedios de los indicadores eficacia y pérdidas, superan el máximo establecido (cinco frutos; Bustillo, 2002).

La situación anterior es posible consecuencia de la falta de capacitación de recolectores, de su falta de conciencia en la labor, de la experiencia de los patrones de corte y de la carencia de metodologías que permitan evaluar en forma rápida y confiable la recolección de café, respecto a los indicadores eficacia y pérdidas.

Hincapié (2005), encontró que de las fincas empresariales de la zona central cafetera que evalúan los indicadores eficacia y pérdidas, aproximadamente el 50% de las fincas realizan la evaluación seleccionando un número de árboles entre 10 o 15 por hectárea y midiendo las correspondientes variables. En el 50% restante de las fincas, se evalúa por criterio visual.

Existe entonces la necesidad de cuantificar el número de frutos dejados en el árbol y en el suelo después de una jornada de recolección, mediante procesos que tengan validez científica.

1.2 FORMULACIÓN

Con el fin de proponer una metodología que sea práctica, se consideró el estudio de dos recorridos para seleccionar árboles en un cafetal.

La pregunta clave a resolver en este trabajo es:
¿Cuál es el recorrido que puede recomendarse para evaluar los indicadores de recolección eficacia y pérdidas?

1.3 SISTEMATIZACIÓN

- ¿Qué diseño de muestreo probabilístico es el indicado para la evaluación de los indicadores eficacia y pérdidas, en una jornada de recolección?

- ¿Cuál o cuáles recorridos proponer para una evaluación rápida?

- ¿Cuáles son las ventajas y desventajas de realizar un muestreo probabilístico con respecto a realizar un recorrido, en la evaluación de los indicadores eficacia y pérdidas?

- ¿Cuál es la significancia de las hipótesis "los valores de los indicadores eficacia y pérdidas, obtenidos por el recorrido propuesto son equivalentes estadísticamente a los obtenidos por el muestreo probabilístico", para cada uno de los recorridos?

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Proponer una metodología para la evaluación de los indicadores eficacia y pérdidas asociados a la recolección de café.

2.2 OBJETIVOS ESPECÍFICOS

- Determinar un diseño de muestra probabilístico para la evaluación de los indicadores eficacia y pérdidas.

- Proponer metodologías alternativas para la evaluación de estos indicadores basados en su carácter práctico en la selección de la muestra.

- Determinar las ventajas y desventajas del muestreo estrictamente aleatorio, como de los recorridos propuestos, en la evaluación de los indicadores eficacia y pérdidas.

- Establecer la significancia de la diferencia entre los valores obtenidos por el diseño de muestra probabilístico y cada recorrido propuesto.

3. JUSTIFICACIÓN

Teórica.

En este trabajo se desea, mediante la aplicación de la teoría y los conceptos básicos de muestreo y las pruebas de aleatorización, proponer una metodología para la evaluación en campo de los indicadores eficacia y pérdidas en una jornada de recolección.

Metodológica.

Para lograr el cumplimiento del objetivo propuesto, se utilizó como base un diseño de muestras probabilístico, que sirvió como estándar para compararse con otro tipo de selección de unidades, mediante una prueba de aleatorización.

Las razones de no proponer un muestreo probabilístico como metodología de evaluación de los indicadores eficacia y pérdidas, se basa en la dificultad de su aplicación en campo y en la necesidad de obtener la información de manera rápida para la consecuente toma de decisiones.

Por ello se propone la evaluación de dos recorridos en el que a priori no se conoce su sesgo ni sus bondades, pero su aplicación requiere de un tiempo menor.

Práctica.

El resultado de este trabajo es la proposición de un método para evaluar los indicadores eficacia y pérdidas en una jornada de recolección, es decir, una herramienta en el control del proceso. El conocer los resultados del proceso, constituye una base para la toma de decisiones, que permitirían mejorarlo, si los resultados de la evaluación así lo sugieren.

4. ANTECEDENTES

4.1 RECOMENDACIONES EN LA RECOLECCIÓN DE CAFÉ

Las recolecciones de café bien realizadas ofrecen ventajas tales como aumentar los ingresos debido a la venta de mayor cantidad de café, reducir las poblaciones de broca y por tanto, evitar futuras reinfestaciones y finalmente, evitar pérdidas hasta del 10%, debidas a los frutos no recolectados o que caen al suelo (Cenicafé, 1998).

Duque (2002), resalta la importancia de "Cosechar solo frutos maduros", como una de las prácticas que apuntan a reducir los costos unitarios de producción o costos por arroba de café pergamino seco producida, mejorando la productividad del cultivo. Todo lo anterior con el fin de que el café sea una empresa competitiva, productiva, sostenible y de cambio técnico. Realizando esta labor de manera selectiva se garantiza una composición adecuada de frutos que permite lograr las mejores conversiones posibles de café cereza a café pergamino seco. Además, el café verde recolectado, se convertirá en pasilla luego del beneficio, y su precio será menor que el café de mejor calidad, y aquellos frutos que se dejen de recolectar caerán al suelo y constituirán en pérdidas económicas para el caficultor. El autor concluye que para productividades de 120 arrobas de café pergamino por hectárea/año, un 10% de pérdida por mala recolección implica dejar de recolectar 12 arrobas por Ha, que equivaldría a $328.000[1]. En el caso de 200 arrobas el

[1] Cifras estimadas para el año 2002.

mismo porcentaje equivale a 20 arrobas por Ha, que tienen un valor aproximado de $548.000/Ha.

Montoya (2001), concluye que al aumentar el porcentaje de frutos maduros desprendidos del árbol, y al recolectar solo frutos maduros, se obtiene un mayor peso de la masa cosechada, una mejor calidad y por lo tanto un menor factor de conversión, y como consecuencia un mayor ingreso. Al dejar menos cantidad de frutos en el suelo, se tiene mayor cantidad de frutos en la masa cosechada, lo cual aumenta el ingreso y facilita el manejo de la broca, disminuyéndose los costos de esta labor.

Federacafé (2000), propone dejar después de un pase, máximo 10 frutos por árbol, sumados los del suelo y los maduros que quedan en las ramas, como una de las prácticas que apunta al control cultural de la broca ya que, la recolección adecuada del fruto, es responsable del 80% del control de esta plaga (Díaz y Marín, 1999). De manera similar, Bustillo (2002), afirma que las labores agronómicas del cultivo, especialmente la cosecha, desempeñan un papel importante en la reducción de las poblaciones de la broca. Esto ha sido comprobado en estudios sobre escape del insecto en el beneficio, en donde se demuestra que entre el 64 y 75% de la población llega al beneficio durante la cosecha. Como indicador de una buena recolección, se ha establecido que solo se debe dejar un máximo de 5 frutos maduros por árbol después de un pase de cosecha.

Duque (2003) plantea que aunque la idea de cero pérdidas en recolección es difícil de alcanzar en los cafetales colombianos, debido a los costos que se generarían, se espera que cualquier acción que el caficultor emprenda con el propósito de reducir las pérdidas en cosecha y cosechar solo frutos maduros, represente un mayor

ingreso bruto por hectárea. Sin embargo, concluye que alguna de las recomendaciones planteadas en este sentido no son económicamente muy atractivas para los recolectores, y es de esperar que las prácticas que no ofrezcan ventajas económicas para ellos, no sean adoptadas. Por ello se deben identificar las limitaciones o ventajas económicas de algunas recomendaciones relacionadas con la cosecha.

Puerta (2000), concluye que para contenidos superiores al 2,5% en peso de café verde en la masa cosechada, beneficiada tradicionalmente o con módulo Becolsub, se deteriora la calidad en taza y el rendimiento del café. Esto implica que en Colombia, es necesaria la realización de múltiples pases, hasta 15 en la región central cafetera (Arcila *et al.*, 2001).

Vélez *et al.* (2003), afirman que al realizarse correctamente la recolección de café, se puede llegar a obtener un producto con las mejores características físicas y organolépticas, que además de asegurar la calidad de la bebida, determinan un mejor precio de venta y la reducción de las pérdidas en campo, debido a la menor cantidad de frutos maduros que quedan después de la cosecha en los árboles y en el suelo.

Desde un punto de vista administrativo, es importante disponer de una base de datos de la recolección, ya que puede orientar al caficultor sobre el desempeño de los recolectores y de la evaluación del proceso en si (Vélez, 2000). Castaño y Jaramillo (s. p. e.), definen la administración de la cosecha de café como el procedimiento que consiste en determinar el momento para realizar los pases de recolección, de tal manera que se pueda determinar la cantidad de recolectores requeridos. Lo anterior se consigue mediante el manejo de registros de florescencias y el manejo de la información de la producción en la finca. Martínez

(2004), indica que las políticas de la finca, el control y supervisión por parte del patrón de corte determinan en gran medida el rendimiento y la calidad en la ejecución de la recolección, lo que repercute directamente en los indicadores del proceso.

4.2 INDICADORES PARA LA EVALUACIÓN DEL PROCESO DE RECOLECCIÓN DE CAFÉ

La medición del logro de los objetivos y de los procesos y sus características es una tarea básica dentro de un proceso (Latorre, 1996). Esta medición se obtiene mediante unos instrumentos llamados indicadores. Un indicador es una variable con la cual se conocen las características de un objeto de trabajo y del grado en el cual se logra un objetivo, asociado con un sistema o proyecto. Los indicadores deben tener las siguientes características (Latorre, 1996):

a. En lo posible deben ser cuantitativos.

b. Deben ser factibles de medir, en términos del tiempo y los recursos que se necesiten para llevar a cabo la medición.

c. Deben medir realmente la forma como se está logrando el cumplimiento del objetivo.

No siempre es posible que un objetivo sea medible a través de un solo indicador, por lo tanto es viable que se presenten varios indicadores que permitan ver diferentes aspectos referentes al logro de un objetivo (Latorre 1996). Los indicadores tienen debilidades y por ello en algunas ocasiones se deben

complementar con otros o estandarizar, para interpretar de manera adecuada lo que tratan de medir. Existen ocasiones en las cuales los indicadores no son independientes entre sí, sino que son complementarios o miden distintos aspectos de un mismo fenómeno.

Los indicadores se constituyen en una fuente de información que, integrada a los procesos organizacionales de toma de decisiones, pueden favorecer al desempeño de las organizaciones, al contribuir a la eficacia de los actores del mismo y a la mejor asignación y distribución de los recursos humanos, materiales y financieros (OCTI, 2004).

Vélez *et al.* (1999), realizaron un estudio en el que definieron las variables que permiten evaluar de una forma adecuada el proceso de recolección de café. De un listado de variables, se escogieron cuatro, cada una correspondiente a un indicador. Esta selección se realizó mediante un análisis multivariado de componentes principales.

Los cuatro indicadores propuestos y las variables que en ese momento se asociaron a cada uno de ellos fueron, respectivamente:

1. Calidad: porcentaje de frutos maduros del total de la masa cosechada.

2. Eficiencia: kilogramos cosechados por unidad de tiempo.

3. Eficacia: porcentaje de frutos maduros desprendidos del árbol, sobre el 100% del total de frutos maduros disponibles para ser recolectados.

4. Pérdidas: porcentaje de frutos dejados en el suelo, sobre el 100% del total de frutos a recolectar en el árbol.

El proceso de recolección se puede evaluar con estos 4 indicadores, y con base en ellos, se pueden optimizar los recursos involucrados en él (Vélez *et al*., 1999).

Villegas (2003), modificó las variables de tres indicadores, de la siguiente manera:

- Calidad: porcentaje de frutos verdes en la masa cosechada.

- Eficacia: número de frutos maduros sin desprender después de la recolección.

- Pérdidas: número de frutos verdes dejados en el suelo.

Las modificaciones se realizaron con los objetivos de:

1. Facilitar el proceso de medición de los indicadores, en el caso de los indicadores eficacia y pérdidas.

2. Facilitar la comparación de los valores obtenidos con los estándares establecidos para el manejo integrado de la broca (MIB) en el caso de eficacia y pérdidas (Bustillo, 2002), y para garantizar las características organolépticas de la bebida, en el caso de calidad (Puerta, 2000).

4.3 ESTUDIOS SOBRE LOS FRUTOS MADUROS DEJADOS EN EL ÁRBOL Y LOS FRUTOS DEJADOS EN EL SUELO

Chamorro *et al.* (1995), evaluaron diferentes prácticas para efectuar los repases de los árboles y las recolecciones de los frutos del suelo, desde el punto de vista económico. El análisis financiero se efectuó mediante la comparación del costo de la mano de obra empleada en cada práctica adicional a la recolección normal, con el ingreso obtenido por el café recuperado en cada una de ellas. Se concluye que resulta factible económicamente, la recolección normal del café cada 15 días para no incurrir en mayores costos y que a pesar de realizar una cosecha cuidadosa, quedan en el árbol y suelo un 10% de los frutos de la cosecha.

Arenas *et al.* (1997), encontraron en 10 municipios del departamento de Caldas que un 47,9% de las fincas realizaban una mala recolección, 28,92% regular y 23,78% buena. Los caficultores dejaron en promedio 16 frutos por árbol de los cuales 5 frutos tenían brocas vivas. También concluyeron que la incidencia de la broca está directamente relacionada con la mayor o menor cantidad de frutos dejados en el árbol. Las variables de índole técnico que influyeron en el aumento de la cantidad de frutos dejados por árbol, son: edad del cultivo, poda del cafeto, variedad de café plantada y modalidad de recolección al contrato. Los caficultores sostienen que la falta de cuidado y de conciencia del recolector en su labor, (interesándole solo sus ingresos), constituyen la causa de los problemas de recolección.

Díaz y Marín (1999) estimaron la cantidad de frutos dejados en la recolección, durante un ciclo productivo del cultivo, clasificando las fincas cafeteras de los municipios de Marquetalia y Palestina (Caldas), en tres categorías: fincas

pequeñas, con áreas entre 1 y 3 ha; fincas medianas, con áreas entre 5 y 8 ha; y fincas grandes, con áreas mayor a 10 ha (Tabla 1).

Tabla 1. Promedios de los frutos maduros dejados en el árbol y frutos en el suelo sin recolectar, y @CPS/ha/año para tres categorías de fincas cafeteras en dos municipios de Caldas.

Categoría	Promedios	Desviación estándar	@ CPS/ha/año
Pequeñas	5 frutos maduros	2,4	10,9
	4 frutos suelo	1,9	
Medianas	8 frutos maduros	5,38	32,8
	4 frutos suelo	3,19	
Grandes	13 frutos maduros	5,63	64,0
	5 frutos suelo	5,66	

Fuente: Díaz y Marín (1999).

En esta investigación se plantearon tres formas de tomar información, para estimar la cantidad de frutos de café cereza no recolectados por árbol, después de la recolección (Díaz y Marín, 1999): 30 árboles seleccionados aleatoriamente, evaluados en 5 grupos de 6 árboles (Método A), 30 árboles en zig-zag (Método B) y 10 árboles seleccionados aleatoriamente (Método C). Ellas se aplicaron en fincas de cada una de las categorías (pequeñas, medianas y grandes), para un total de 9 fincas por municipio y 18 fincas en total para la investigación. Con base en los promedios que tuvieron la menor desviación estándar, se propuso para Marquetalia, adoptar el método C, mientras que para Palestina se recomendó el método A. Para cada forma, no se presentó ni el error de estimación, ni la probabilidad asociada a él.

Villegas (2003), concluyó que la altura de la planta no incide en el comportamiento del número de frutos dejados en el árbol (eficacia), número de frutos dejados en el suelo (pérdidas), ni en el porcentaje de frutos verdes en la masa cosechada (calidad), pero si en el indicador de eficiencia (kilogramos recolectados por hora), de tal manera que, en la menor altura evaluada, 157,7 cm, los recolectores aumentaron su eficiencia. Durante las primeras cuatro jornadas de trabajo, la eficacia y las pérdidas por operario fueron más variables que en las siguientes jornadas (5-10) (Tabla 2).

La calidad evaluada como el porcentaje de frutos verdes en la masa cosechada, según lo obtenido por Villegas (2003), tuvo coeficientes de variación entre 34,5 y 67,6%, con promedios por debajo del 2,1% (Tabla 3), de tal manera que en dicha investigación se obtuvo una calidad que no altera la bebida, según el estándar propuesto por Puerta (2000).

Tabla 2. Promedios y coeficientes de variación para los indicadores de recolección.

INDICADORES DE RECOLECCION	**Jornada de trabajo (días)**	**Prom.**	**C.V. (%)**
Eficacia (No. de frutos maduros sin desprender)	1 – 4	12,4 A	99,8
	5 – 10	5,6 B	29,0
Calidad (% frutos verdes en la masa cosechada)	1 – 4	1,2 A	40,6
	5 – 10	0,9 B	44,8
Pérdidas (No. de frutos dejados en el suelo)	1 – 4	21,5 A	61,2
	5 – 10	10,8 B	29,3

Fuente: Villegas (2003).

Tabla 3. Promedio y coeficiente de variación del indicador Calidad, para cada operario.

	Calidad (% de frutos verdes en la masa cosechada)	
Operario	Prom.	C. V. (%)
1	1,0	52,1
3	1,2	34,5
4	1,0	61,3
5	1,1	67,6
6	1,4	36,3
7	2,1	33,6
8	1,8	42,0
9	1,0	43,9
10	2,0	35,4

Fuente: Villegas (2003).

Villegas (2003) también estimó los promedios para los indicadores eficiencia, calidad y pérdidas, para cada operario, en árboles de distinta altura máxima (Tabla 4). Los mayores coeficientes de variación y sus respectivos promedios, para las variables asociadas a los indicadores eficacia y pérdidas fueron 91,1%, 9,7 frutos; y 77,0%, 11,8 respectivamente. Las varianzas asociadas a estos promedios son 78,1 y 82,6 para eficacia y pérdidas respectivamente. De manera similar a la tabla anterior, no se presentaron problemas respecto al porcentaje de frutos verdes en la masa cosechada, solo un promedio superó el valor 2,5% (operario 7, altura 203,3 cm).

Tabla 4. Promedios y coeficientes de variación para los indicadores eficacia y pérdidas por árbol, y calidad, para cada operario, en árboles de diferentes alturas.

INDICADOR		Eficacia		Pérdidas		Calidad	
Operario	Altura promedio (cm)	Prom.	C.V (%)	Prom.	C.V (%)	Prom.	C.V (%)
1	164,6	6,8	57,1	8,8	66,9	1,4	57,6
	157,7	6,0	51,3	9,6	75,1	0,7	31,1
	203,3	9,7	91,1	10,7	61,4	1,0	41,8
	239,5	8,1	56,4	8,4	34,7	0,9	44,9
3	164,6	8,1	40,6	10,8	63,0	1,3	20,9
	157,7	6,6	43,0	13,4	62,3	1,3	28,3
	203,3	11,8	46,2	14,5	62,6	1,2	57,2
	239,5	8,9	64,6	7,8	24,9	0,9	14,5
4	164,6	9,2	37,0	8,8	51,3	0,9	69,6
	157,7	9,7	54,1	13,5	73,4	0,9	32,2
	203,3	10,9	46,4	15,2	68,5	1,1	35,2
	239,5	7,6	49,9	7,2	62,9	1,3	86,7
5	164,6	8,9	43,9	11,0	34,6	1,5	85,8
	157,7	10,5	34,0	11,0	61,3	1,4	28,8
	203,3	9,4	73,7	13,9	58,8	1,0	29,4
	239,5	10,9	38,7	8,0	52,8	0,7	48,0
6	164,6	8,9	53,4	11,3	55,0	1,3	43,4
	157,7	11,3	18,7	15,9	49,3	1,3	23,6
	203,3	11,6	63,0	18,6	42,0	1,5	38,8
	239,5	17,5	31,5	17,4	33,8	1,6	44,0
7	164,6	10,1	50,3	13,1	46,7	1,8	20,5
	157,7	12,4	30,8	17,3	41,6	1,9	19,9
	203,3	15,4	59,8	18,2	43,9	2,9	31,6
	239,5	13,0	29,8	15,8	27,6	1,7	25,0
8	164,6	14,1	31,8	14,0	37,6	2,0	49,0
	157,7	15,6	31,8	16,2	51,7	1,6	65,1
	203,3	15,1	21,6	15,7	52,6	2,0	38,5
	239,5	13,3	21,0	14,3	51,8	1,7	21,5
9	164,6	8,1	52,5	9,7	26,6	0,9	42,2
	157,7	9,9	36,2	14,1	55,4	0,8	7,3
	203,3	11,9	60,8	13,4	32,1	1,2	55,0
	239,5	9 1	34,2	14,4	40,5	1,0	47,7
10	164,6	9,9	44,5	15,1	41,9	2,1	54,2
	157,7	10,5	19,6	11,8	77,0	1,9	39,8
	203,3	15,8	49,3	21,8	32,1	2,0	28,9
	239,5	14,0	31,6	17,1	24,0	2,1	27,4

Fuente: Villegas (2003)

4.4 MÉTODOS EXISTENTES EN LA EVALUACIÓN DE RECOLECCIÓN

En el 66% de las fincas empresariales, de la zona central cafetera, se realiza un control sobre el número de frutos maduros sin recolectar después de un pase (Hincapié, 2005). Sólo en el 38%, se evalúan los frutos dejados en el suelo. Al menos en el 50% de las fincas, la evaluación de los frutos dejados en el árbol y suelo se realiza con criterio visual. En el resto de fincas, se seleccionan un número de árboles que oscila de 10 a 15, y en ellos se cuentan los frutos dejados. Esta recomendación fue enunciada por Cenicafé (1995).

Respecto al porcentaje de frutos verdes en la masa cosechada, la evaluación se realiza en el 81% de estas fincas, de tal manera que el 69% lo hace por observación y el 21% toma una muestra de 1 kg, del cual pesa la cantidad del fruto verde.

5. MARCO TEÓRICO

Los fundamentos teóricos en este trabajo tuvieron como base los siguientes aspectos: Planeación de un muestreo, y Pruebas de aleatorización.

5.1 PLANEACIÓN DE UN MUESTREO

Consiste en definir una serie de etapas que se deben realizar para la ejecución de una investigación de muestreo. A continuación se describen cada una de esas etapas.

- **Definición de objetivos.** Deben ser claros y concisos y su desarrollo se cumple realizando el muestreo (Scheaffer *et al*., 1987).

- **Población objetivo.** Se entiende por población un agregado de datos individuales, personas o cosas, acerca de las cuales se desea información (Chou, 1975). Los elementos individuales de una población se llaman unidades elementales. Definir una población es limitar el contenido de las unidades elementales. Estas poseen ciertas características que pueden ser de naturaleza cualitativa o cuantitativa. El resultado de una observación de una unidad elemental se llama *dato*.

La población debe ser definida cuidadosa y claramente en términos de los elementos que la componen de tal forma, que se pueda establecer si un elemento

pertenece o no a ella; también se debe definir de una manera tal que la selección de una muestra de sus elementos sea realmente factible.

- **Marco muestral.** Es el medio físico donde se listan los elementos de la población objetivo de estudio. El marco de muestreo está conformado por las unidades de muestreo (Scheaffer *et al*., 1987). Las *unidades de muestreo* son colecciones no traslapadas de elementos de la población que tienen una cobertura completa sobre ella. Si cada unidad de muestreo contiene uno y solamente un elemento de la población, entonces una unidad de muestreo y un elemento de la población son idénticos. La lista de las unidades de muestreo, conforman el marco de muestreo.

- **Variables.** La definición de las variables de la investigación debe especificar la naturaleza de las características, las reglas de la categoría de clasificación, y las unidades para expresarlas. Deben determinar también el alcance y el contenido del marco muestral (Kish, 1979). Una variable puede ser de dos tipos:

1. ***Cualitativa*** o ***categórica***, si los elementos pueden ser clasificados de acuerdo a sus valores como poseedores o no, de cierta cualidad o propiedad (Chou, 1975). Una variable cualitativa se puede clasificar en dos escalas (Behar y Yepes, 1996): ***nominal***, si hace uso de números para clasificar los elementos en distintos grupos, clases o categorías de acuerdo con alguna propiedad cualitativa; y ***ordinal***, si hace uso de los números no solo para clasificar los elementos de un conjunto en categorías, sino para realizar comparaciones de la forma: "más grande", "igual", "menor", es decir, el valor numérico de la medida se usa para ordenar las categorías de mayor a menor de acuerdo a un criterio.
2. ***Cuantitativa***: es aquella variable expresada en datos numéricos simplemente por medición directa en la unidad elemental (Chou, 1975). Las mediciones en las

unidades elementales forman la totalidad de las observaciones, se expresan numéricamente en una unidad de medición determinada, y conforman los *valores* que una variable puede asumir. Una variable cuantitativa puede medirse en dos escalas: ***continua***, si puede asumir cualquier valor numérico (número real) dentro de una amplitud específica, en la que las unidades pueden dividirse en fracciones de cualquier tamaño, por pequeñas que sean, de modo que haya un flujo continuo de tales valores con graduaciones infinitamente pequeñas; y ***discreta***, si solo puede adoptar ciertos valores, generalmente enteros.

- **Parámetros.** Un parámetro es la medida obtenida de observaciones en toda la población (Chou, 1975). Un parámetro es un solo valor, que describe en forma resumida las características pertinentes o un estado de naturaleza acerca de una población. Aún cuando el muestreo se realiza para múltiples propósitos, el interés se centra frecuentemente en cuatro parámetros (Cochran, 1978): media ($\overline{Y}$ o μ) , total (Y o τ), proporción (P) y razón (R).

- **Diseño de muestreo.** El método de selección de una muestra, se denomina *diseño de muestreo*. El diseño debe incluir el número de elementos en la muestra y la manera como se van a seleccionar (Scheaffer *et al*., 1987).

Los diseños de muestreo se clasifican en *probabilístico* o *no probabilístico*. Elegir entre alguno de las dos, depende de los objetivos del estudio, del esquema de investigación y de la contribución que se piensa hacer con ella (Hernández *et al*, 1998).

1. Muestreo no probabilístico. La característica principal en este tipo de muestreo es que las unidades son seleccionadas de acuerdo a un criterio por parte del investigador, es decir, no hay un método objetivo por el que se prefiera un criterio

de selección de muestra a otro (Pérez, 2000). Se caracteriza porque los elementos de la muestra no tienen una probabilidad conocida de selección (Méndez, 2001). Además, el error de muestreo no puede ser medido, ya que su principal limitante es que no se puede utilizar la inferencia estadística para obtener conclusiones objetivas y válidas para toda la población (Fracica, 1998).

Pérez (2000) clasifica los tipos de muestreo no probabilístico en:

a) Muestreo intencional u opinático: la persona selecciona la muestra de acuerdo a su experiencia en el área de investigación, juzgando así la representatividad de la selección realizada. También se le conoce como muestreo de experto.

b) Muestreo por cuotas: se selecciona a las personas de manera no aleatoria de acuerdo con una cuota determinada (Fracica, 1998). Existen dos tipos de muestreo de cuota: proporcional, cuando se conoce las proporciones en que se encuentra dividida la población respecto a una categoría de interés; y no proporcional, donde no se conocen estas proporciones, pero se especifica el número mínimo de unidades de muestra que se requiere en cada categoría.

c) Muestreo sin norma o por conveniencia: se selecciona la muestra de cualquier manera por razones de comodidad. La representatividad de la muestra solo puede aspirar a ser medianamente satisfactoria en el caso de que la población sea muy homogénea. Este tipo de muestreo suele aplicarse en la vida corriente, siempre que en caso de equivocación las consecuencias no sean demasiado graves. Se utilizan cuando se necesitan estimaciones a partir de las cuales no se piensa tomar una decisión importante, siendo a su vez el presupuesto de la investigación pequeño.

d) Muestreo semiprobabilístico: puede ser de dos clases: superior, cuando se conoce la probabilidad de selección de cierta parte de la población, más no la de un elemento dentro de ella; e inferior cuando se conoce la probabilidad de selección de los elementos dentro de la parte o segmento, más no la del segmento mismo.

2. Muestreo probabilístico: en este tipo de muestreo, cada elemento de la población tiene una probabilidad conocida de ser seleccionado en una muestra (Scheaffer *et al.*, 1987). En la selección de la muestra, se toma en cuenta le teoría de los números aleatorios, es decir, se realiza de una manera objetiva, la cual es la principal diferencia con los diseños de muestreo no probabilístico.

En este tipo de muestreo, el diseño y el tamaño de la muestra determinan la cantidad de información pertinente a un estimador de un parámetro poblacional, es decir, una estimación (puntual o intervalo), la varianza del estimador, el error de estimación y la probabilidad asociada a él.

Los cuatro diseños de muestreo probabilístico básico son:

a) Muestreo Aleatorio Simple.

b) Muestreo Aleatorio Estratificado.

c) Muestreo Sistemático.

d) Muestreo por conglomerados.

A continuación se presentan las principales características de ellos.

a) Muestreo Aleatorio Simple. Se realiza una selección aleatoria de *n* unidades, en una población de tamaño *N*, donde todas las unidades elementales tienen la misma probabilidad conocida de ser seleccionadas (Scheaffer *et al.*, 1987). Los elementos son enumerados y son seleccionados si están asociados a un número aleatorio en la muestra. En general, cuando un elemento ya ha sido seleccionado, no se toma en cuenta de nuevo para la selección de los siguientes, es decir se trata de un muestreo sin reemplazo. En otras ocasiones se utiliza el muestreo con reemplazo, es decir una unidad seleccionada es tomada en cuenta de nuevo para la selección de la próxima unidad.

Constituye la base de los otros diseños de muestreo probabilístico, y por las propiedades que posee con relación a la estimación de los parámetros y los errores de muestreo, se ha convertido en el método más utilizado (Klínger, 2003).

b) Muestreo Aleatorio Estratificado. En ocasiones, la población que se desea estudiar está compuesta de subgrupos bien definidos que pueden ser identificados con anterioridad, y cuyos elementos son mutuamente excluyentes. Estos subgrupos son llamados estratos (Ospina, 1992). El muestreo aleatorio estratificado se realiza mediante un muestreo aleatorio simple dentro de cada estrato, y las estimaciones para los parámetros de la población total se obtendrán combinando las estimaciones para los estratos.

La principal característica de los estratos, es la homogeneidad de las unidades dentro de ellos, y la heterogeneidad entre estratos. Una estratificación de la población se puede realizar después de la obtención de los datos, lo cual es conocido como post-estratificación.

Las razones principales del uso del muestreo aleatorio estratificado son (Scheaffer *et al*., 1987):

1. La estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por un muestreo aleatorio simple, con un mismo tamaño de muestra, si las unidades dentro de los estratos son homogéneas, reduciendo el costo por observación de cada unidad.

2. Se pueden obtener estimaciones de parámetros poblacionales para subgrupos de la población.

c) Muestreo Sistemático. En este método los elementos de la población se listan en una secuencia ordenada y la selección de la muestra se hace con base en un intervalo constante en esta lista (marco muestral). Los elementos de la población son marcados con los números de 1 a N. El tamaño de la población se divide por el tamaño de la muestra *n* para determinar el intervalo muestral *k*, que se define como el entero más próximo a *N/n*. Se selecciona aleatoriamente un número entre 1 y *k,* el cual se denomina *r*, y corresponde al primer elemento de la muestra, a partir del cual se realiza la selección cada *k* elementos.

Un estimador por muestreo aleatorio simple y uno de muestreo sistemático, son igualmente eficientes cuando la *población es aleatoria* (Scheaffer *et al*., 1987); cuando la *población es ordenada,* es decir los elementos dentro de la población están ordenados en magnitud de acuerdo con algún esquema, un muestreo sistemático proporciona mayor información que un muestreo aleatorio simple, es decir es mas eficiente; y cuando la *población es periódica (*sus elementos tienen variación cíclica), con un muestreo aleatorio simple se obtiene mayor eficiencia que

con un muestreo sistemático, ya que proporciona mayor información por unidad de costo (Scheaffer *et al*., 1987).

La utilidad del muestreo sistemático, se visualiza de una manera más clara debido a que, por su facilidad en lo que hace referencia a la selección de la muestra en el campo, está menos expuesto a los errores de selección que cometen los investigadores de campo por lo tanto cuando realizan muestras aleatorias simples y estratificadas, (Scheaffer *et al*., 1987).

d) Muestreo por Conglomerados. En este tipo de muestreo la unidad de muestreo consiste en un grupo o conglomerado de unidades más pequeñas (Cochran, 1978) y no se necesita tener un marco dentro de cada conglomerado, solo una lista de los conglomerados es necesaria para realizar este diseño de muestreo. Se selecciona un grupo de conglomerados usando muestreo aleatorio simple o estratificado, y dentro de los conglomerados seleccionados hay dos opciones: el proceso de medición se realiza en todas las unidades, o se realiza un muestreo aleatorio simple dentro de cada conglomerado seleccionado, y se realiza la medición en las unidades seleccionadas.

La facilidad que representa la selección de conglomerados, y que redunda en un costo menor, conlleva a estimadores menos eficientes; es importante anotar entonces, que este diseño es efectivo para obtener una cantidad especificada de información al costo mínimo, siempre y cuando exista homogeneidad en cuanto a las medias de la variable entre ellos, y alta variabilidad dentro de cada conglomerado. Esta última es una condición necesaria para poder aprovechar las ventajas económicas del muestreo por conglomerado (Scheaffer *et al*., 1987).

- **Estimadores.** Es una función que depende de todas los valores obtenidos en la medición de las unidades seleccionadas en una muestra, y su objetivo es estimar el valor verdadero de un parámetro (Pérez, 2000). El estimador es una variable aleatoria, al considerar la variabilidad de la selección de una muestra. El valor obtenido mediante el uso de un estimador en una muestra se denomina estimación. La estimación resultante puede ser acompañada por un límite para el error de estimación o expresada como un intervalo de confianza.

La precisión de un estimador se cuantifica mediante su varianza, su desviación estándar o su error cuadrático medio (Pérez, 2000). Los estimadores de los parámetros se simbolizan de la siguiente manera: $\hat{\mu}$ o $\hat{Y}$, media; $\hat{\tau}$ o $\bar{y}$, total; $\hat{p}$, proporción; y $\hat{r}$, razón.

Un aspecto importante al realizar el plan de muestreo es conocer la distribución de probabilidad de los estimadores que se utilizarán, para así poder evaluar la magnitud del error de estimación, es decir el valor absoluto de la diferencia entre el parámetro y el estimador. Para conocer estas distribuciones se debe tener la distribución de la población objetivo de estudio.

A continuación se ilustran los estimadores del diseño de muestreo aleatorio simple, presentados por Scheaffer *et al*. (1987), para los parámetros media, total y proporción (Tabla 5).

En la tabla 5, s^2 representa la varianza muestral o $s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$, y n es el número de unidades seleccionadas en la muestra.

Tabla 5. Estimadores y varianzas respectivas, para los parámetros media, total y proporción en el muestreo aleatorio simple.

Parámetro	Estimador	Varianza del estimador
Media	$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$	$\hat{V}(\bar{y}) = \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$
Total	$\hat{\tau} = N\bar{y} = \frac{N\sum_{i=1}^{n} y_i}{n}$	$\hat{V}(\hat{\tau}) = N^2 \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$
Proporción	$\hat{p} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$	$\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$

- **Error de estimación.** El error de estimación es el valor absoluto de la diferencia entre el parámetro y la estimación. No se asegura que un estimador observado estará dentro de una distancia especificada de θ (Scheaffer *et al.*, 1987), pero se puede al menos encontrar un límite aproximado B tal que

$$P(\left|\hat{\theta} - \theta\right| \leq B) = 1 - \alpha$$

para cualquier probabilidad deseada $1 - \alpha$, donde $0 < \alpha < 1$. Esta probabilidad es conocida como el *coeficiente de confianza*. Si $\hat{\theta}$ tiene una distribución normal, entonces $B = z_{\alpha}\sigma_{\hat{\theta}}$, donde z_{α} es el valor para el cual la función de distribución acumulada normal es $1 - \alpha/2$, y $\sigma_{\hat{\theta}}$ es la desviación estándar conocida o estimada para el estimador.

Existen diversos estimadores cuya distribución no es exactamente una normal, para muchos valores de n (tamaño de la muestra) y N (tamaño de la población). El Teorema de Tchebysheff establece que al menos 75% de los valores para cualquier distribución de probabilidad están dentro de dos desviaciones estándar de su

media, por lo cual usualmente se usa $2\sigma_{\hat{\theta}}$ como un límite para el error de estimación (Scheaffer *et al*., 1987); cuando se fija este límite de error, los coeficientes de confianza son del 95 % para estimadores con distribución normal, y de al menos 75% para estimadores con otra distribución.

- **Tamaño de muestra.** El tamaño de muestra depende principalmente del conocimiento previo de la varianza muestral, y el establecimiento del error de estimación y su respectivo coeficiente de confianza por parte del investigador, en cualquiera de los diseños de muestreo probabilístico. Para la estimación de la varianza, se recurre a investigaciones anteriores, y si no se dispone de información previa, se debe realizar una prueba piloto.

Scheaffer *et al*. (1987), presenta las fórmulas de los tamaños de muestra para la estimación de los parámetros media, total y proporción, para cada diseño de muestreo probabilístico, de acuerdo a una varianza estimada, un error establecido igual a dos desviaciones estándar del estimador, y un coeficiente de confianza de al menos el 75%. Las fórmulas para hallar el tamaño de muestra en el muestreo aleatorio simple se presentan en la tabla 6.

En la tabla 6, B_μ y B_T son los errores establecidos para los estimadores de la media y del total, respectivamente. B_p es el error establecido para el estimador de proporción.

- **Organización del trabajo de campo.** Se definen las unidades a medir, las variables, el método de medición, y la forma en realizar las mediciones (Pérez, 2000). Para ello, el personal involucrado en el proceso de medición en las unidades elementales debe ser entrenado y supervisado (Scheaffer *et al*., 1987).

Tabla 6. Tamaños de muestra para la estimación de la media, el total y la proporción en el muestreo aleatorio simple.

Parámetro	Tamaño de muestra (*n*)	D
Media	$\frac{Ns^2}{(N-1)D+s^2}$	$\frac{B_\mu^2}{4}$
Total	$\frac{Ns^2}{(N-1)D+s^2}$	$\frac{B_\tau^2}{4N^2}$
Proporción	$\frac{N\hat{p}(1-\hat{p})}{(N-1)D+\hat{p}(1-\hat{p})}$	$\frac{B_p^2}{4}$

- **Análisis de los datos.** Se definen los análisis que se deben realizar a la información recolectada, para establecer las estimaciones finales y sus variaciones (Scheaffer *et al.*, 1987). Adicionalmente, se evalúan los errores de estimación obtenidos y se plantean estrategias y/o ajustes para próximos estudios similares (Klínger, 2003).

5.2 PRUEBAS DE ALEATORIZACIÓN

Una prueba estadística para la cual la significancia de los resultados experimentales es determinada permutando los datos repetidamente para calcular *t*, *F*, o algún otro estadístico de prueba, es llamada una prueba de aleatorización (Edgington, 1995). Las pruebas de aleatorización, cuando se calculan los estadísticos de prueba convencionales, no son *alternativas* a las pruebas convencionales; más bien, ellas son pruebas donde la significancia se determina por un procedimiento especial.

En cada reordenamiento de los datos se calcula un estadístico de prueba. El conjunto resultante de estadísticos de prueba en la totalidad de reordenamientos realizados constituye la distribución muestral (a menudo llamada distribución de referencia) de ese estadístico. Se puede usar esa distribución de referencia para determinar la significancia de la hipótesis nula, que es el porcentaje de los reordenamientos que tienen un valor tan extremo o más al valor obtenido con los datos experimentales sin permutar o reordenar (Howell, 2002).

Esta es una manera de obtener la significancia, diferente a la de las pruebas paramétricas. Para realizar estas últimas, es necesario extraer muestras aleatoriamente de una o más poblaciones. Se deben cumplir ciertos supuestos poblacionales, esencialmente, la normalidad y la igualdad de varianzas entre ellas. Se establece una hipótesis nula que es construida en términos de parámetros, por ejemplo, la comparación de las medias de dos poblaciones, H_0: $\mu_1 - \mu_2 = 0$. Se usa el valor del estadístico tanto para la estimación, como para calcular un estadístico de prueba. Se compara el valor del estadístico con una distribución muestral del estimador, y se rechaza la hipótesis nula si el estadístico de prueba es relativamente extremo en la distribución.

Las pruebas de aleatorización difieren de las pruebas paramétricas en los siguientes aspectos (Howell, 2002).

- No existe el supuesto de que se tienen muestras aleatorias de una o más poblaciones. De hecho, usualmente no se han extraído las muestras aleatoriamente.

- Rara vez, se piensa en términos de las poblaciones de las cuales vienen los datos, y no hay necesidad de asumir nada acerca de normalidad u homocedasticidad.

- La hipótesis nula generalmente no se refiere a parámetros. Se expresa de una manera precisa, diciendo que, bajo la hipótesis nula, el valor asociado a un individuo es independiente del tratamiento que ese individuo recibió, si es un experimento, o el grupo al cual pertenece ese individuo, si es una investigación no experimental.

- Como no se toma en cuenta la concepción de alguna población, el objetivo no es realizar pruebas sobre los parámetros de ellas.

- Se calcula alguna clase de estadístico de prueba, pero no se compara este estadístico con una tabla de distribuciones establecida, sino con la distribución de referencia, constituida por los valores que toma el estadístico de prueba en cada aleatorización de los datos originales. Es decir, se construye la distribución de referencia del estadístico de prueba, para así estimar el valor *p*.

Las pruebas paramétricas de todas las clases, incluyendo pruebas relativamente complejas como análisis de varianza factorial, análisis de covarianza, y análisis multivariado de varianza, se vuelven libres de distribución cuando la significancia es determinada por una prueba de aleatorización. Dada la asignación aleatoria, un investigador puede garantizar la validez de una prueba estadística usando una prueba de aleatorización para determinar la significancia (Edgington, 1995).

La derivación original de la distribución muestral de un nuevo estadístico de prueba y la formulación de una tabla de significancia, lleva un tiempo considerable

y demanda un alto nivel de habilidad matemática, si un investigador estableciese una nueva prueba estadística que sirviera en su caso particular. Sin embargo, si el investigador determina la significancia mediante las permutaciones, no se necesita mucho tiempo o habilidad matemática, debido a que la "tabla de significancia" es generada por permutaciones de los datos.

La versatilidad de las pruebas de aleatorización tiene base en que se puede establecer un estadístico de prueba, de tal manera que incremente la probabilidad de un investigador de conseguir resultados significantes cuando existen diferencias reales en los efectos de los tratamientos. La posibilidad de asegurar la validez de cualquier prueba estadística existente, proporciona al investigador la oportunidad de seleccionar la prueba que sea más sensible a una diferencia entre grupos, o tratamientos si es el caso. Además, el investigador no está limitado a pruebas estadísticas existentes; el uso de las pruebas de aleatorización permite el desarrollo de pruebas estadísticas que podrían ser más sensibles a los efectos de tratamiento, comparadas con las pruebas convencionales, y estas nuevas pruebas estadísticas pueden ser desarrolladas para acomodar radicalmente diferentes procedimientos de asignación aleatoria (Edgington, 1995).

Algunas de las múltiples aplicaciones de las pruebas de aleatorización son: comparación de dos o más muestras independientes, análisis de varianza de medidas repetidas, diseños factoriales, y análisis de varianza con un factor de clasificación entre otros (Edgington, 1995; Congedo, 2000; Howell, 2002). Los investigadores no deberían encontrar dificultad en generalizar a otras pruebas estadísticas.

Procedimiento.

La manera de realizar una prueba de aleatorización es (Howell, 2002):

1. Se define y se calcula un estadístico de prueba para los datos originales.

2. Se realizan *Q* permutaciones o reordenamientos de los datos, de tal manera que se mezclen la totalidad de ellos entre los grupos.

3. Para cada permutación, se calcula el mismo estadístico de prueba calculado para los datos originales. La totalidad de los valores de ellos, constituirán el conjunto de referencia para determinar la significancia.

4. La proporción de las permutaciones en el conjunto de referencia que tengan valores del estadístico de prueba mayores o iguales que (para ciertos estadísticos de prueba, menores o iguales que) el estadístico de prueba de los resultados obtenidos experimentalmente es el valor *P* (significancia o valor de probabilidad).

5. Se rechaza o se retiene la hipótesis nula con base en esta probabilidad.

Este procedimiento puede ser realizado con cualquier prueba de aleatorización. Como se mencionó anteriormente, se pueden realizar diversas pruebas reordenando aleatoriamente los datos, pero cada prueba determina tanto la manera de realizar el proceso de aleatorización, y un estadístico de prueba que es el más sensible para cada prueba específica (Edgington, 1995).

Edgington (1995) divide las pruebas de aleatorización en dos, según el método de construir la distribución de referencia del estadístico: en las que se utilizan método

sistemático, es decir toman en cuenta todas las posibles combinaciones de los datos; y las que se toman con una muestra aleatoria de todas las posibles combinaciones, llamado método aleatorio.

Algunos autores denominan las pruebas de aleatorización, como pruebas de permutación, aunque sería más preciso llamarlas como combinaciones, antes que permutaciones. Sin embargo, el término pruebas de permutación es aceptado en la literatura. Howell (2002), afirma que en algunas ocasiones considerar todas las posibles combinaciones es imposible, como en el caso de tener tres grupos con 20 observaciones cada uno, se obtendrían 0,7783 x 10^{26} combinaciones, y aún la supercomputadora más rápida no es capaz de extraer todas esas muestras. Esto se podría realizar si realmente fuese necesario, pero ciertamente no valdría la pena. Este autor propone tomar una muestra aleatoria de todas las posibles combinaciones. La respuesta obtenida no será exacta, pero será tan cercana que no habrá alguna diferencia en la decisión tomada. Los resultados de 5000 muestras ciertamente serán tan cercanos a la respuesta exacta, que la diferencia vendrá desde el tercer lugar decimal o más allá.

El extraer muestras aleatorias para estimar los resultados de todas las muestras posibles, es un caso particular de la aplicación de un proceso de simulación Monte Carlo. Se realiza principalmente debido a que el tiempo y el dinero para realizar todas las permutaciones o combinaciones posibles para los datos es demasiado grande. Se determina la significancia como la proporción de los valores del estadístico de prueba que son tan extremos como el obtenido, para el número de permutaciones realizadas. Este es un procedimiento válido para determinar la significancia, que puede ser demostrado inmediatamente. Bajo H_0, el valor del estadístico de prueba obtenido puede ser considerado como aleatoriamente seleccionado del conjunto de todos los posibles valores del estadístico de prueba,

así que se tienen Q valores seleccionados aleatoriamente, uno de los cuales es el valor obtenido. En general, cuando H_0 es cierta, la probabilidad de alcanzar un valor P tan pequeño como P no es mayor a P, y así cumple el criterio de validez general presentado por Edgington (1995):

Criterio de validez general. *Un procedimiento de prueba estadística es válido, si, bajo la hipótesis nula, la probabilidad de un valor de significancia tan pequeño como P no es más grande que P, para cualquier P.*

Por ejemplo, bajo la hipótesis nula la probabilidad de obtener un valor de significancia (valor p) tan pequeño como 0,05 tiene que ser no mayor a 0,05; un valor de significancia de 0,03 tiene que ser no mayor a 0,03, y así sucesivamente.

Una prueba de aleatorización que use el método de permutación aleatorio, aunque emplee únicamente una muestra de todas las posibles permutaciones, es por consiguiente válida. Cuando H_0 es verdadera, la probabilidad de rechazarla a cualquier nivel α no es mayor que α; pero si H_0 es falsa y existe una verdadera diferencia entre grupos, o un efecto real de tratamientos, la permutación aleatoria es menos potente que la permutación sistemática basada en el conjunto de todas las asignaciones posibles. Al incrementar el número de permutaciones usadas con el método de permutaciones aleatorias, se incrementa la potencia de una prueba de aleatorización empleando permutación aleatoria.

Validez en las pruebas de aleatorización.

Existe un numeroso grupo de investigadores quienes están interesados en usar la pequeñez de un valor P como un indicador de la fuerza de la evidencia contra H_0

(una interpretación de valores *P* inconsistente con la aproximación por la teoría de la decisión).

La validez de las pruebas de aleatorización para varios tipos de aplicación ha sido demostrada por diversos autores. Sin embargo, el uso de un procedimiento por si solo no garantiza la validez. A la luz de numerosos estadísticos de prueba y procedimientos de asignación aleatoria que pueden ser realizados en las pruebas de aleatorización, es esencial para el investigador saber las reglas básicas para la ejecución válida de una prueba de aleatorización. Estas reglas básicas son útiles también a los investigadores que no establecen niveles de significancia de antemano pero en cambio usan la pequeñez del valor *P* como un indicador de la fuerza de la evidencia contra H_0. Es importante conocer las condiciones que tienen que ser satisfechas para ciertos procedimientos de prueba para conocer el criterio de validez general. Dependiendo del procedimiento, las condiciones o supuestos podrían incluir muestreo aleatorio, asignación aleatoria, normalidad, homogeneidad de varianza, homocedasticidad, u otras varias condiciones (Edgington, 1995).

¿Por qué se consideran equiprobables las selecciones de las muestras?

Las selecciones pueden ser consideradas equiprobables debido a la hipótesis nula, la cual establece que los valores de X_1 y los valores de X_2, si se trata de dos muestras, son todos independientes e idénticamente distribuidos (Conover, 1999). Por ello los valores de X_1 no deberían tener una tendencia definida a ser más bajos mientras que los valores de X_2, fuesen más altos, o viceversa. Dado cualquier grupo de $m + n$ números, si ellos son observaciones o no, cada subgrupo de n de esos números es tan probable a que correspondan a los n valores de X_1 como cualquier otro grupo de n de esos números, debido a que los números que no

pertenecen a X_1, tienen que pertenecer a X_2, y la probabilidad asociada a ese grupo de números no depende de cuales números son llamados X_1 y cuales son llamados X_2 sino que depende de que ambos provienen de una misma población, esto con el objetivo de encontrar una región crítica de tamaño a. Este es el argumento intuitivo para considerar cada selección de n observaciones de X_1s como igualmente probables (Conover, 1999).

La manera para construir la distribución de referencia en estas técnicas, es similar a la desarrollada por los autores de la estadística no paramétrica para las pruebas basadas en rangos (Edgington, 1995; Conover, 1999).

Para una prueba de aleatorización, el supuesto principal es "intercambiabilidad". Esto se reduce básicamente al supuesto de que si la hipótesis nula fuese cierta, las asignaciones de las observaciones de los sujetos a los grupos son intercambiables (Howell, 2002). En otras palabras, si dos grupos no fuesen diferentes, o un tratamiento no tuviese efecto (en el caso de un diseño experimental), un individuo tendría el mismo valor, sin importar a que grupo pertenece, o a que tratamiento fue asignado. Por ello, aun después de que los datos han sido recolectados, la media o cualquier medida estadística de lo que se ha llamado "grupo uno", tendría la misma probabilidad de salir después de haberse mezclado los sujetos entre los grupos. Esta es la razón de crear una distribución de referencia (distribución de muestreo) tomando los datos a mano, mezclándolos, y reasignándolos a los grupos. Cuando se realiza este proceso repetidamente, los resultados que se obtendrán serán la distribución de referencia para la hipótesis nula.

Muestreo aleatorio y asignación aleatoria.

El principal supuesto de los procedimientos paramétricos es que se ha realizado una muestra aleatoria de una población. En algunas situaciones no tiene sentido, estadísticamente, decir que las medias muestrales son estimaciones de los parámetros poblacionales correspondientes a menos que las muestras sean extraídas aleatoriamente de esas poblaciones, por ejemplo al estimar el efecto de un medicamento. Algunas veces el supuesto de aleatoriedad es crítico. Si no se muestrea aleatoriamente, no se pueden realizar inferencias significativas acerca de los parámetros, siguiendo los lineamientos de la inferencia estadística clásica.

Una prueba de aleatorización es válida para cualquier clase de muestra, indiferente de cómo fue seleccionada. Esta es una propiedad extremadamente importante debido a que el uso de muestras no aleatorias es común en la experimentación, y las tablas estadísticas paramétricas (por ejemplo, tablas *t* y *F*) no son válidas para tales muestras (Edgington, 1995).

Los argumentos respecto a la "representatividad" de una muestra seleccionada no aleatoriamente son irrelevantes a la cuestión de su aleatoriedad: una muestra aleatoria es aleatoria debido al procedimiento de muestreo usado para seleccionarla, no debido a la composición de la muestra. Es decir, la aleatoriedad, no conlleva a la representatividad. La selección aleatoria es necesaria para asegurar que las muestras sean aleatorias, y se cumpla uno de los supuestos fundamentales para el uso de las tablas estadística paramétricas.

Las violaciones del supuesto del muestreo aleatorio pueden ser variadas. Una violación es usar todos los sujetos disponibles, sin emplear un procedimiento de muestreo de cualquier clase. Un segundo tipo de violación es la selección

sistemática de cierta clase de sujetos. Una tercera forma de violación puede ocurrir aún, cuando se ha realizado un muestreo aleatorio: realizando inferencias estadísticas acerca de una población sobre la base de muestras aleatorias de otra población, tales como una subpoblación. Por consiguiente es esencial para la validez de las tablas paramétricas estadísticas asegurar que no solamente se tiene que una población ha sido muestreada aleatoriamente sino que también la población muestreada es sobre la cual se realizarán inferencias estadísticas.

Por supuesto, si se tomarán muestras aleatorias de las poblaciones sobre las cuales se desean realizar inferencias estadísticas, las tablas de significancia serían conocidas, y se podrían validar o no los supuestos paramétricos relevantes (tales como normalidad y homogeneidad de varianza) concernientes a las poblaciones. Al ser validados, las tablas estadísticas paramétricas pudiesen ser empleadas validamente para determinar la significancia (Edgington, 1995).

El enfoque de las pruebas de aleatorización a un problema es totalmente diferente. En primer lugar, no se le da a una "población" la importancia que ella recibe en las pruebas paramétricas. No se piensa en la forma de su distribución, ni en la idea de extraer muestras aleatorias de esas poblaciones imaginarias. No existe preocupación en estimar los parámetros, por lo que **no se necesitan muestras aleatorias**. Solo se desea conocer la verosimilitud de los resultados obtenidos si los grupos no tuviesen diferencias. Por ello, no se necesita pensar absolutamente sobre poblaciones (Howell, 2002).

En la historia de la estadística, los procedimientos más conocidos fueron desarrollados con base en una muestra aleatoria, debido a que la idea de la estimación era el punto central.

Cotton (1973, p. 168)[2] anota que las pruebas de aleatorización permitirían validar inferencias estadísticas en la ausencia de muestreo aleatorio. Él continuó la idea de Fisher al enfatizar que determinar la significancia por medio de las tablas estadísticas convencionales, es de validez cuestionable hasta que se haya mostrado que la significancia obtenida es similar a la que se obtendría mediante una prueba de aleatorización.

Spino y Pagano (1991)[3] se refirieron a un seminario de Tukey (1988)[4], llamado "Aleatorización y re-aleatorización: la onda del pasado en el futuro", en el cual él reveló su continua creencia de que la aplicación de las pruebas de aleatorización, a la cual él se refiere como "re-aleatorización", provee los mejores resultados para experimentos aleatorizados en general. Si al usar las tablas de significancia siempre se obtuviesen aproximaciones cercanas a los valores de significancia dados por las pruebas de aleatorización, las pruebas de aleatorización, que son los únicos procedimientos completamente válidos de usar para encontrar la significancia con muestras no aleatorias, serían únicamente de interés académico, más no de aplicación. Sin embargo, las tablas de significancia, algunas veces proporcionan valores que difieren considerablemente de aquellos dados por las pruebas de aleatorización, y así el asunto de la validez es prácticamente un problema. Por supuesto, también existen ocasiones en las cuales los valores *P* dados por los dos métodos de obtener la significancia son muy similares. En casos de grandes diferencias, la diferencia es a veces en una dirección y a veces en otra. Varios factores afectan la proximidad de la significancia proporcionada por estos dos métodos de obtener significancia para permitir la formulación de una regla

[2] COTTON, J. W. Even better than before. En: Contemp. Psychol. No. 18 (1973): p. 168–169.

[3] SPINO, C. y PAGANO, M. Efficient calculation of the permutation distribution of trimmed means. En: J. Am. Statist. Assn. No. 86 (1991); p. 729–737.

[4] TUKEY, J. W. Randomization and rerandomization: the wave of the past in the future. En: Seminar given in honor of Joseph Ciminiera. Philadelphia Chapter of the American Statistical Association, June, 1988.

general para el investigador que le indique el tamaño del error que se cometería al usar las tablas paramétricas en lugar de las pruebas de aleatorización. Algunos de estos factores son el tamaño de muestra absoluto, los tamaños de muestra relativos para los diferentes grupos o tratamientos, la forma de la distribución muestral, el número de medidas repetidas, y el tipo de prueba estadística (Edgington, 1995).

A pesar de la conocida importancia de las pruebas de aleatorización, no se espera que ellas reemplacen las tablas de significancia como el medio más común para determinar la significancia. La razón es que, por el uso rutinario de las pruebas comúnmente usadas como la prueba *t* y el ANOVA de una vía, la validez de las tablas de significancia rara vez es puesta en tela de juicio. En las aplicaciones menos comunes de las pruebas estadísticas, surgen inquietudes sobre la validez, y es por tales situaciones que la determinación de la significancia por una prueba de aleatorización es especialmente valiosa. La aplicación de un ANOVA a datos dicotómicos, por ejemplo, probablemente despierta dudas sobre su validez en situaciones donde la aplicación a datos cuantitativos no se utilizara. El problema de la validez también nace si un experimentador desea aplicar un ANOVA a datos de un experimento con un solo individuo. Con las pruebas menos usadas comúnmente, como el análisis de covarianza y el análisis multivariado de varianza, cualquier clase de aplicación podría generar discusiones sobre el sustento de los supuestos fundamentales para utilizar las tablas estadísticas. Ciertamente, cuando primero se introduce una nueva prueba estadística, su validez debe ser demostrada. Como consecuencia de las múltiples maneras en las cuales la validez de los valores *P* puede ser cuestionada, no es inusual que un investigador utilice una prueba estadística menos potente que la que él tiene en mente, o aun evitar experimentos de cierta clase, para tener resultados publicables. Se pueden evitar

estos contratiempos usando las pruebas de aleatorización, en vez de las tablas estadísticas convencionales (Edgington, 1995).

Al usar las pruebas de aleatorización, las muestras probablemente no han sido extraídas aleatoriamente, y se asume que no tienen (ni necesitan) supuestos válidos para estimar parámetros poblacionales. Esto conlleva a que tampoco hay bases válidas para tomar decisiones sobre hipótesis nulas sobre esos parámetros, que es la base de cualquier prueba estadística. En la prueba de aleatorización, se tiene una manera alternativa para probar estas hipótesis, aunque estas no sean hipótesis nulas, en el sentido estricto de la palabra (Howell, 2002).

El tema de la aleatoriedad de la muestra no es la única diferencia entre las pruebas de aleatorización y las pruebas paramétricas. En estas últimas existe gran preocupación sobre la asignación aleatoria, debido que el principio de asignación aleatoria dice que si la hipótesis nula es cierta, se podrían mezclar o reordenar los datos y esperar a conseguir exactamente los mismos resultados. Esto es por lo que la asignación aleatoria es fundamental en el procedimiento estadístico empleado. Sin embargo, existen situaciones en los que esta asignación no se puede realizar, por ejemplo, un caso con niños esquizofrénicos, donde se trabajaría con grupos establecidos.

Los defensores de la prueba de aleatorización como Edgington y Lunneborg no se han preocupado mucho sobre los supuestos sobre las varianzas o la normalidad de las poblaciones. Estas ideas no son centrales en la lógica que servirá para responder la pregunta clave. Las personas a favor de las pruebas de aleatorización sienten orgullo al afirmar que no deben asumir homogeneidad de varianza, pero a menudo olvidan que esto es debido a que ellos deben responder una pregunta diferente (Howell, 2002). En las pruebas paramétricas se preguntan,

por ejemplo, si las medias son diferentes (parámetros), mientras que en las pruebas de aleatorización la cuestión es si los tratamientos o grupos tienen "efectos" diferentes, y no se usa el término "efecto" en el sentido estadístico técnico del diseño experimental. No debería convertirse en sorpresa que cuando se plantee una pregunta diferente, se deban realizar diferentes supuestos, y se obtendrán respuestas diferentes.

Tradicionalmente, las pruebas de aleatorización han sido representadas como pruebas de hipótesis nulas de diferencias en los grupos o en los efectos de tratamiento; sin embargo, se pueden probar otras hipótesis, como por ejemplo, la hipótesis nula respecto a cierta magnitud de efecto, o la hipótesis de relaciones cuantitativas (Edgington, 1995).

La ausencia de asignación aleatoria.

Las pruebas de aleatorización se basan en la idea de la asignación aleatoria para justificar la permutación de los datos, de acuerdo a la hipótesis nula. Sin embargo, existen algunas situaciones donde los grupos de estudio ya han sido establecidos, es decir sus individuos no han sido asignados aleatoriamente. Existe entonces una contradicción entre el supuesto de asignación aleatoria y el interés de determinar si se presentan diferencias entre estos grupos (Howell, 2002).

Howell (2002) no sugiere que la asignación aleatoria no es importante, y que no importa realmente si se asignan los casos a los grupos o tratamientos aleatoriamente. Más bien, afirma que simplemente debido a que no se tiene asignación aleatoria no existe excusa para no realizar un análisis a los datos, lamentándose que no hay nada que se pueda hacer. Es mucho lo que se puede hacer, aún si las conclusiones finales son menos precisas de lo que nos gustaría

que ellas fueran. Este es un punto importante, debido a que muchas de las comparaciones que nos gustaría realizar, son entre grupos pre-existentes para los cuales la asignación aleatoria es imposible.

Estadísticos de prueba equivalentes.

Dos estadísticos de prueba que proporcionen el mismo valor *P* para una prueba de aleatorización son llamados estadísticos de prueba equivalentes (Edgington, 1995). Esta idea es de notable importancia, en la tarea de reducir el tiempo de corrida de un algoritmo en computador, al establecer un estadístico de prueba que es más fácil de calcular que otro que es equivalente a él. Dos estadísticos de prueba son equivalentes si y solo si están perfectamente correlacionados de manera monótona a través de todas las permutaciones de los datos. Expresado en términos de correlación, será una correlación positiva o negativa entre los valores de los dos estadísticos de prueba través del conjunto de permutaciones usado para determinar significancia. El primer paso es simplificar un poco un estadístico de prueba paramétrico convencional, y cada paso sucesivo simplifica el estadístico de prueba resultante del paso anterior. El estadístico de prueba para una prueba de aleatorización puede ser más simple debido que necesita reflejar únicamente la diferencia entre grupos, o el efecto de tratamiento de interés, a diferencia de los estadísticos de prueba paramétricos, los cuales deben involucrar también una estimación de la variabilidad del componente reflejando la diferencia de los grupos o el efecto de tratamientos. Por ejemplo, el numerador de *t*, el cual es la diferencia entre medias, es sensible al efecto de tratamientos; el denominador es necesario para estimar la variabilidad del numerador. Para otros estadísticos de prueba paramétricos también, el denominador frecuentemente sirve para estimar la variabilidad del numerador. En una prueba de aleatorización. bajo la hipótesis nula, el procedimiento de la prueba genera su propia distribución de la medida de

la diferencia de grupos o tratamientos, tal como una diferencia entre medias, haciendo irrelevante al denominador (Edgington, 1995).

Pruebas de aleatorización cuando se presenta una variable de clasificación.

En la búsqueda bibliográfica también se encontró que las pruebas de aleatorización se pueden realizar cuando se presenta un factor de confusión o de clasificación de los datos. En algunas situaciones, puede existir un factor *T* que constituya un sistema de clasificación (Steel y Torrie, 1988), y se desea determinar si existe diferencia entre de los grupos (factor *A*) sin tener en cuenta la clasificación de los datos. Los reordenamientos aleatorios tienen la restricción de realizarse dentro de los valores que tome esta variable, no entre ellos. Esto es similar al proceso de aleatorización en un diseño de bloques generalizado (Díaz-Uriarte, 2001). Para este diseño, el estadístico de prueba se calcula para los datos, teniendo en cuenta el factor de clasificación, en este caso el bloque, es decir este es el *F* de un modelo de diseño de dos vías (Díaz-Uriarte, 2001). La distribución de referencia se construye de manera similar a como se ha enunciado anteriormente para los otros casos, al igual que el cálculo del valor *p*.

Edgington (1995), menciona que un estadístico de prueba equivalente y más fácil de manejar al momento de realizar los cálculos para un análisis de varianza con un factor de clasificación, es $\Sigma(T_i^2/n_i)$, siendo T_i el total de las observaciones en un grupo, y n_i es el número de observaciones en ese grupo. En cada permutación, se realiza el proceso de aleatorización o mezcla dentro de cada valor de la variable de clasificación, y se calcula ese estadístico de prueba. Sin embargo, se presenta un problema con este estadístico cuando los tamaños de las muestras entre los grupos no son proporcionales. En una base de datos, los tamaños de muestra son

completamente proporcionales si la razón de los tamaños de muestra de dos celdas cualesquiera en una fila es la misma tal como la razón de los tamaños de muestra para las celdas correspondientes en otra fila. Un estadístico de prueba alternativo, es la razón de la varianza entre los grupos a comparar, intervarianza, y la varianza dentro de cada uno de estos grupos, intravarianza, principio fundamental del análisis de varianza. El objetivo principal de un análisis de varianza es determinar si dos variables son independientes. En términos estadísticos, dos variables X y Y, son independientes si la distribución de Y es la misma en todos los subgrupos que se construyan con base en la variable X (Behar y Yepes, 1996).

Cuando se hace referencia a la variabilidad de la media de Y, en los diferentes subconjuntos o valores de la variables X, es decir, la varianza de $M(Y/x_1)$, $M(Y/x_2)$, …, $M(Y/x_m)$ se habla de la intervarianza, que se denota por S_E^2.

De igual forma existe interés de conocer la magnitud de la varianza de Y al interior de cada subconjunto de datos, es decir cuando se quiere tener una idea sobre la magnitud de las varianzas: $S_{Y/x1}^2$, $S_{Y/x2}^2$,…, $S_{Y/xm}^2$. El promedio de estas varianzas, se conoce como intravarianza y se simboliza S_D^2.

Se puede establecer que a medida que haya más diferencias entre los grupos, el valor de la intervarianza será mayor. Un estadístico de prueba para probar si los grupos difieren es el cociente entre la inter y la intra varianza (Behar y Yepes, 1996).

Sin embargo, como se mencionó al referirse a los estadísticos de prueba equivalentes, el procedimiento de la prueba de aleatorización genera la distribución, bajo la hipótesis nula, de la medida de la diferencia de grupos o

tratamientos, haciendo irrelevante al denominador donde se estima la intravarianza.

Como resultado de lo anterior, el estadístico de prueba para comparar dos grupos, cuando se presenta un factor de clasificación es la suma de cuadrados entre grupos, una función lineal de la intervarianza (Edgington, 1995).

Procedimiento para realizar la prueba de aleatorización con un factor de clasificación.

1. Se organizan los datos como en la tabla 7. Se desea conocer si entre los valores de la variable *A* existen diferencias en la respuesta sin tener en cuenta el factor de clasificación (*T*). No existe restricción sobre igualdad en el número de observaciones en cada celda.

2. Estadístico de prueba. En el caso donde se presenta un factor de clasificación, Edgington (1995) propone el siguiente estadístico de prueba cuando se presentan tamaños de celdas no proporcionales:

$$S_E^2 = \sum_{i=1}^{a} n_{i.}[\bar{y}_{i.} - E(\bar{y}_{i.})]^2$$

siendo $E(\bar{y}_{i.}) = \frac{\sum_{j=1}^{t} n_{ij}\bar{y}_{.j}}{n_{i.}}$, una media ponderada de cada grupo, que depende de la media dentro de el valor j de la variable de clasificación, y el número de unidades en cada celda.

Este estadístico S_E^2 es una generalización de la suma de cuadrados entre grupos en un análisis de varianza. La modificación consiste en sustituir la gran media general por las medias esperadas para cada grupo. Esta sustitución no tiene

efecto cuando hay tamaños de celda proporcionales, pero si tiene un efecto considerable cuando los tamaños de celda no son proporcionales, haciendo la prueba más sensible.

Tabla 7. Observaciones y_{ijm} (*a*) y número de observaciones por celda (*b*) n_{ij} ´s.

	(*a*)						(*b*)					
Niveles de la variable *A*	Niveles de la variable *T*				Total	Media	Niveles de la variable *A*	Niveles de la variable *T*			Total
	1	2	…	*t*				1	2	…	*T*	
1	y_{111} . . . $y_{11n(11)}$	y_{121} . . . $y_{12n(12)}$	…	y_{1t1} . . . $Y_{1tn(1t)}$	$y_{1..}$	$\overline{y}_{1.}$	1	n_{11}	n_{12}	…	n_{1t}	$n_{1.}$
	$Y_{11.}$	$y_{12.}$	…	$y_{1t.}$								
2	y_{211} . . . $y_{21n(21)}$	y_{221} . . . $y_{22n(22)}$	…	y_{2t1} . . . $y_{2tn(2t)}$	$y_{2..}$	$\overline{y}_{2.}$	2	n_{21}	n_{22}	…	n_{2t}	$n_{2.}$
	$Y_{21.}$	$y_{22.}$	…	$y_{2t.}$								
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A	y_{a11} . . . $y_{a1n(a1)}$	y_{a21} . . . $y_{a2n(a2)}$	…	y_{at1} . . . $y_{atn(at)}$	$y_{a..}$	$\overline{y}_{a.}$	*A*	n_{a1}	n_{a2}	…	n_{at}	$n_{a.}$
	$y_{a1.}$	$y_{a2.}$	…	$y_{at.}$								
Total	$y_{.1.}$	$y_{.2.}$	…	$y_{.t.}$	$y_{...}$		Total	$n_{.1}$	$n_{.2}$	…	$n_{.t}$	$n_{..}$
Media	$\overline{y}_{.1.}$	$\overline{y}_{.2.}$	…	$\overline{y}_{.t.}$		$\overline{y}_{...}$						

3. Se realizan los Q reordenamientos aleatorios dentro de cada nivel de clasificación j. Para cada una de las Q tablas resultantes de estos reordenamientos, se calcula el estadístico de prueba S_E^2.

4. Se determina el porcentaje de los reordenamientos aleatorios, o tablas resultantes, cuyo valor del estadístico de prueba es mayor al obtenido para los datos originales, es decir, cuando no se han permutado las observaciones.

5. Este porcentaje es el valor p estimado. Si es menor a la significancia establecida antes de la prueba, se rechaza la hipótesis nula.

Historia de las pruebas de aleatorización.

Edgington (1995) presenta un resumen sobre los avances de la técnica de las pruebas de aleatorización, y el aporte realizado a ella por diversos autores.

André (1883)[5] desarrolló una fórmula recursiva para realizar una prueba de rachas, generando una distribución de números de rachas (secuencias de valores ascendentes o descendentes) para $M!$ ordenamientos de una secuencia de N valores numéricos. La distribución de probabilidad resultante puede ser usada para realizar una prueba de permutación sobre el supuesto del muestreo y/o de la asignación aleatoria. Como el número de rachas depende únicamente de los rangos de los valores numéricos y no de los valores reales, ha sido posible proporcionar una tabla para esta prueba de aleatorización, la cual es presentada en Bradley (1968, p. 363)[6].

[5] ANDRÉ, D. Sur le nombre de permutations de *n* éléments qui présentent *s* séquences. En: Comptes Rendus (Paris). No. 97 (1883); p. 1536-1358.

[6] BRADLEY, J. V. Distribution-free Statistical Tests. New Jersey: Prentice Hall, Engl. Cliffs, 1968.

Fisher (1935, sección 21)[7] llevó a cabo una prueba de aleatorización, en la cual empleó un conjunto de referencia de valores de un estadístico de prueba, que eran dependientes, no solo de los valores de los rangos, sino de los valores reales de las mediciones involucradas. Aparentemente, esta fue la primera prueba de permutación que empleó un conjunto de referencia que había sido derivado separadamente para cada conjunto diferente de datos; a diferencia de las pruebas de aleatorización anteriores a esta, las cuales utilizaron únicamente las propiedades ordinales de los datos, esta prueba no permitió la derivación de tablas de significancia generalizadas. El experimento hipotético y la prueba fueron descritas por Fisher de la siguiente forma: Cincuenta semillas fueron seleccionadas aleatoriamente de cada una de dos poblaciones para probar la hipótesis nula de que las poblaciones son idénticas respecto al tamaño de las plantas que las semillas producirían bajo las mismas condiciones. Las semillas podrían no haber crecido bajo condiciones idénticas, y para prevenir diferencias poblacionales al ser confundidas con las diferencias ambientales y del suelo, las semillas en las muestras fueron asignadas aleatoriamente a las macetas en las cuales serían plantadas. Las macetas con suelo y ubicación similares fueron emparejadas, y para cada uno de los 15 pares de macetas se determinó aleatoriamente cual unidad de un par de semillas sería plantada en cada maceta. Después de que las plantas alcanzaran cierta edad, se midieron las alturas, y cada par de semillas asignadas a macetas con igual condición ambiental, se emparejaron, con el fin de realizar una prueba *t* apareada. Fisher consideró el conjunto referencia de $2^{15} = 32768$ permutaciones asociadas con el cambio de signos de las diferencias para todas las 15 parejas, y calculó la diferencia entre los totales para los dos tipos de semillas como estadístico de prueba. El valor *P* fue la proporción de esas 32768 permutaciones con una diferencia tan extrema entre los totales como la obtenida con los datos originales. Sin la ayuda de un computador, Fisher calculó el valor *P*.

[7] FISHER, R. A. Design of experiments. Edinburgh: Oliver and Boyd, 1935.

Esa difícil tarea, aún para Fisher, es reflejada en su necesidad de revisar su primer cálculo del valor *P* en las últimas ediciones de su libro "Diseño de experimentos"[8].

Frecuentemente, la prueba de Fisher ha sido discutida como si solo estuviese basada en la asignación aleatoria, como lo están las pruebas de aleatorización modernas. Por ejemplo, una serie de documentos de seis autores sobre "la prueba de aleatorización de Fisher" encabezada por un artículo de Basu (1980)[9], discutió la prueba mencionada como si en ella no estuviese involucrado el muestreo aleatorio, lo cual es inconsistente con la descripción de Fisher de la prueba. Fisher en muchos puntos de su discusión se refirió a la prueba de hipótesis nulas de identidad de poblaciones, y, de hecho, el muestreo aleatorio de poblaciones fue requerido para la prueba. Sin selección aleatoria de las semillas, la hipótesis nula probada podría haber sido identificar el efecto de las macetas apareadas sobre el crecimiento de las plantas, un efecto que no interesaba a Fisher. La asignación aleatoria en el ejemplo de Fisher fue para controlar las diferencias entre macetas, no para probar estas diferencias. Las unidades experimentales (semillas) no fueron asignadas aleatoriamente con el propósito de determinar el efecto de las macetas a las cuales ellas fueron asignadas, y el cambio del signo en las diferencias entre los totales dentro de las parejas no se realizó para representar lo que habría resultado bajo la hipótesis nula si la asignación de semillas dentro de un par hubiera sido invertida. Fue la hipótesis nula de poblaciones infinitas idénticas la que justificó el cambio del signo de la diferencia entre observaciones apareadas: el muestreo aleatorio de las poblaciones más la asignación aleatoria dentro de las parejas aseguró que cualquier diferencia absoluta entre los totales tuvo un signo de más o de menos con igual probabilidad. Se resalta que la prueba de aleatorización de Fisher, a diferencia de las pruebas de aleatorización

[8] FISHER, R. A. Op. cit.

[9] BASU, D. Randomization analysis of experimental data: the Fisher randomization test. En: J. Am. Statist. Assn. No. 75 (1980); p. 575-582.

modernas, no involucró asignación aleatoria a la prueba del efecto de los tratamientos experimentales ni requirió muestreo aleatorio de las poblaciones infinitas (Edgington, 1995).

Pitman (1937ª, b, 1938)[10] proporcionó la primera estructura teórica para las pruebas de aleatorización y pruebas de permutación no experimentales. En su primer documento sobre pruebas de aleatorización, Pitman (1937ª, p. 119) rechazó la precedencia de su teoría por medio de esta cautelosa afirmación: "La idea principal no es nueva, *parece* estar *implícita* en todos los escritos de Fisher" (se adicionan las cursivas). Pitman (1986)[11] posteriormente hizo este comentario sobre esa afirmación: "Siempre estuve insatisfecho con la frase que escribí… Lo que quise decir realmente es que yo estaba haciendo algo nuevo." Pitman (1937ª, b, 1938) dio varios ejemplos de las pruebas de permutación también como un razonamiento aplicable tanto a las pruebas de aleatorización como a las pruebas de permutación no experimentales. Después de proporcionar un razonamiento para las pruebas de aleatorización de las diferencias entre poblaciones, el mostró que las pruebas de permutaciones podrían estar basadas solo en asignación aleatoria, enfatizando su libertad del supuesto de muestreo aleatorio y dando un ejemplo en cada uno de los tres documentos (Pitman, 1937ª, p. 129; 1937b, p.231; 1938, p. 322). Esta demostración es de gran importancia para un amplio número de investigadores quienes usan la asignación aleatoria de individuos u otras unidades que no han sido seleccionadas aleatoriamente.

---

[10] PITMAN, E. J. G. Significance tests which may be applied to samples from any populations. En: J. R. Statistic. Soc. B. No. 4 (1937a); p. 119-130.

_____. Significance tests which may be applied to samples from any populations. II. The correlation coefficient. En: J. R. Statist. Soc. B. No. 4 (1937b); p. 225-232.

_____. Significance tests which may be applied to samples any populations. III. The analysis of variance test. En: Biometrika. No. 29 (1938); p. 322-335.

[11] CARTA DE E. J. G. Pitman. 1986.

Un número importante de pruebas de rango fueron desarrolladas brevemente después de los artículos de Pitman. Las pruebas de rango tuvieron la tendencia a ser representadas como pruebas de hipótesis sobre poblaciones con base en muestras aleatorias, aún en una investigación experimental. Las pruebas de permutaciones de rangos han sido ampliamente usadas debido a sus tablas de significancia fácilmente disponibles. La aplicabilidad de las pruebas de rango a los datos experimentales con base en solo la asignación aleatoria tendió a ser descuidada, como lo fue la posibilidad de permutar datos originales de manera similar a como son permutados los rangos para tener una prueba más sensible (Edgington, 1995).

Mientras las pruebas de permutación para rangos fueron rápidamente desarrolladas dentro de la estructura del muestreo aleatorio, no había un interés manifestado en las pruebas de permutación basadas en la aleatorización, es decir en las pruebas de aleatorización. El uso principal de las pruebas de aleatorización fue estimular la reevaluación del análisis de varianza y otros procedimientos de la curva normal para el análisis de los datos experimentales. Kempthorne (1952, 1955)[12] y sus colegas, por ejemplo, hicieron un trabajo considerable sobre el "análisis de aleatorización", el cual se refirió a la investigación de las distribuciones de los estadísticos de prueba paramétricos, tanto las determinadas por la hipótesis nulas, como las determinadas por las hipótesis alternativas, (especialmente *F*) bajo la aleatorización experimental. El análisis de aleatorización reflejó la actitud tomada hacia las pruebas de aleatorización por muchos investigadores principiantes: las pruebas de aleatorización serían herramientas útiles para la evaluación de la validez de las pruebas paramétricas bajo la violación de sus

[12] KEMPTHORNE, O. Design and Analysis of Experiments. New York: Wiley, 1952.
_____. The randomization theory of experimental inference. En: J. Am. Statist. Assn., No. 50 (1955); p. 946-947.

supuestos. La liberación de las pruebas de aleatorización de ese rol subordinado, y el reconocimiento del valor intrínseco de ellas, se retrasó hasta la disponibilidad de computadores que las hicieran prácticas y posibles de realizar. Durante muchos años, Kempthorne hizo importantes contribuciones al campo de las pruebas de aleatorización. Él y unos pocos colegas mantuvieron a los investigadores conscientes de la necesidad de examinar detalladamente el vínculo entre los procedimientos de aleatorización y las pruebas estadísticas aplicadas a los resultados experimentales. Kempthorne proporcionó una alternativa al razonamiento de las "permutación" para las pruebas de aleatorización: él describió la generación de lo que el llamó permutaciones realizadas sobreponiendo los resultados experimentales sobre cada uno de los posibles patrones de asignación. Este concepto, llamado referencia de aleatorización, es extremadamente útil para entender el razonamiento de las pruebas de aleatorización.

Las pruebas de aleatorización entraron en su época con la llegada de la era computacional. Los computadores personales ahora pueden generar rápida y económicamente miles de permutaciones para realizar una prueba de aleatorización. No obstante, con solo tamaños de muestra moderados podrían ser posibles muchas asignaciones que no serían factibles de proporcionar, aún para computadores modernos. (Existen por ejemplo, cerca de 5 trillones de maneras de asignar 30 individuos a tres tratamientos con 10 individuos por tratamiento.) Dwass (1957)[13] ayudó a resolver este problema proporcionando un razonamiento para "pruebas de aleatorización modificadas" basadas en muestras aleatorias de todas las posibles permutaciones. El conjunto de referencia generado aleatoriamente, por ejemplo, podría consistir de solo 1000 de esas permutaciones. El procedimiento de la prueba de aleatorización propuesto por Dwass, no es simplemente una aproximación Monte Carlo a una "verdadera prueba de

[13] DWAS, M. Modified randomization tests for nonparametrics hypotheses. En: Ann. Math. Statist. No. 28 (1957); p. 181-187.

aleatorización", sino que es una prueba completamente válida, aún cuando el número de permutaciones en el subconjunto aleatorio es una porción muy pequeña del número asociado con todas las asignaciones posibles. El uso de este procedimiento de permutación aleatorio, **el cual es válido para pruebas de aleatorización no experimentales**, fue facilitado por la construcción de algoritmos computacionales para permutación aleatoria de datos, realizada por Green (1963, pp. 172-173)[14] y otros.

Otro procedimiento diseñado para realizar pruebas de aleatorización de manera más práctica, reduciendo el número de permutaciones en el conjunto de referencia fue propuesto Chung y Fraser (1958)[15]. Su procedimiento también involucró el uso de solamente un subconjunto de las permutaciones asociadas con todos las posibles asignaciones, pero a diferencia del subconjunto de Dwass, el de ellos era un subconjunto seleccionado sistemáticamente (no aleatorio). Es dudoso que la aproximación de Chung y Fraser, basada en el razonamiento de grupo de permutación, sea tan útil como el procedimiento aleatorio para reducir el tiempo computacional para una prueba de aleatorización, pero su concepto de grupo de permutación es de considerable importancia teórica. Su aproximación proporciona una justificación sólida teóricamente para muchas aplicaciones de las pruebas de aleatorización a complejos diseños experimentales, proporcionando por esto una base para una expansión sustancial del alcance de la aplicabilidad de las pruebas de aleatorización.

Solo se han considerado las principales contribuciones, pero es suficiente para indicar que se ha logrado mucho progreso desde el trabajo pionero de Fisher y Pitman en los años 30´s. A medida que la gente tome conciencia del potencial de

[14] GREEN, B. F. Digital Computers in Research. New York: McGraw-Hill, 1963.

[15] CHUNG, J. H. y FRASER, D. A. S. Randomization tests for a multivariate two-sample problem. En: J. Am. Statist. Assn. No. 53 (1958); p. 729-735.

las pruebas de aleatorización, las contribuciones en el campo se volverán más frecuentes.

La prueba como método computacional intensivo.

Los métodos estadísticos computacionales intensivos, los cuales dependen sobre la capacidad de procesamiento de los computadores, incluyen una extensa variedad de pruebas estadísticas y procedimientos de estimación. Algunos procedimientos intensivos computacionales como los métodos de remuestreo calculan valores *P* de las distribuciones de estadísticos de prueba derivados directamente de los datos, aunque no necesariamente de la misma manera que en las pruebas de aleatorización. (Algunos autores a menudo incluyen las pruebas de aleatorización dentro de los procedimientos de remuestreo.) Investigadores, estadísticos, y científicos computacionales han contribuido al desarrollo de las pruebas y procedimientos de estimación basados en el remuestreo, haciéndolos disponibles en varios paquetes comerciales.

Tanto las pruebas de aleatorización experimentales, como las no experimentales, tienen una fuerte base teórica proporcionada por Pitman (1937ª, 1937b, 1938)[16], y desarrollada por autores posteriores. Esas pruebas no solo han sido la base en considerables investigaciones para datos en general, sino también en el desarrollo de pruebas de rango, cuyo razonamiento depende de la teoría de las pruebas de permutación.

En reportes investigativos y en libros no paramétricos el término "pruebas de aleatorización" es a veces usado para referirse a cualquier prueba de permutación, y la aplicación de sus "pruebas de aleatorización" a datos de investigación no

[16] PITMAN, E. J. G. Op. cit.

experimental o individuos seleccionados aleatoriamente es presentada como si el muestreo aleatorio fuese innecesario, como de hecho sería para el caso de una verdadera prueba de aleatorización. Tal práctica pasa por alto una importante propiedad de las pruebas de permutación: el procedimiento computacional tiene que ser suplementado con muestreo u asignación aleatoria con el fin de proporcionar pruebas de hipótesis válidas.

6. METODOLOGIA

6.1 TIPO DE ESTUDIO

La investigación fue exploratoria transeccional. El trabajo de campo fue realizado en dos localidades del departamento de Caldas, en los meses de marzo, abril y mayo del año 2005.

6.2 DISEÑO DE LA INVESTIGACIÓN

Con el fin de poder cumplir los objetivos del trabajo se llevó a cabo el siguiente esquema metodológico:

1. Desarrollo de un plan de muestreo probabilístico para la evaluación de los indicadores eficacia y pérdidas.

2. Desarrollo de dos técnicas de muestreo alternativas para la evaluación de los indicadores eficacia y pérdidas.

3. Recolección de información.

4. Análisis de la información.

6.3 PLAN DE MUESTREO PROBABILÍSTICO PARA LA EVALUACIÓN DE LOS INDICADORES EFICACIA Y PÉRDIDAS

Se realizó un plan de muestreo, basado en un diseño de muestreo probabilístico. Las etapas de este plan fueron: definición del objetivo, la población objetivo de estudio, el marco de muestreo, los parámetros de interés, el diseño de muestreo, los estimadores y sus respectivas varianzas, el error de estimación, la confiabilidad, y el tamaño de muestra.

El objetivo del plan de muestreo fue estimar después de una jornada de recolección, los promedios de las variables asociadas a los indicadores de eficacia y pérdidas por árbol, así como las varianzas de ellas.

La población objetivo de estudio, fueron los árboles recolectados por un grupo de operarios en una jornada. La unidad de muestreo fue cada árbol recolectado en esa jornada. El marco de muestreo correspondió a un mapa realizado antes de cada jornada, donde se identificó cada árbol y cada surco del terreno donde se realizaría el proceso de recolección.

Las variables consideradas en el estudio fueron: número de frutos maduros (incluyendo sobremaduros y secos) dejados en el árbol (eficacia); y número de frutos dejados en el suelo (pérdidas). Los parámetros de interés para ambas variables fueron el promedio y la varianza.

Debido a que no se conocían criterios de estratificación, ni agrupamiento de los frutos dejados en el árbol y en el suelo, el diseño de muestreo probabilístico realizado fue el **aleatorio simple**.

Las fórmulas de los estimadores para los promedios fueron tomadas de Scheaffer *et al*. (1987) para un muestreo aleatorio simple (Tabla 5). Se estimaron los intervalos de confianza ($\bar{y} \pm \hat{\sigma}_{\hat{y}} t_{\alpha/2}$, donde $\hat{\sigma}_{\hat{y}}$ es el error estándar de la media muestral (Tabla 5) y $t_{\alpha/2}$ es el valor en la tabla de la distribución *t* de Student para el cual $P(t > t_{\alpha/2}) = \alpha/2$).

El tamaño de muestra para cada variable, se estableció con base en una varianza estimada, y un error de estimación para el parámetro y su respectivo coeficiente de confianza. Los valores de las varianzas estimadas de ambas variables, fueron tomados de Villegas (2003) (Tabla 4). Las varianzas consideradas fueron 78,1 para el número de frutos maduros en el árbol; y 82,6 para el número de frutos en el suelo. Se tomó esta última con el fin de establecer un tamaño de muestra que cumpliera el error establecido para ambas variables.

Para calcular el tamaño de muestra se utilizó la fórmula propuesta por Scheaffer *et al*. (1987) para la estimación del promedio (Tabla 6). Asumiendo una varianza de 82,6, un error de dos frutos para el promedio, un coeficiente de confianza de este error de al menos 75%, y una población de 10.000 árboles, se encontró un tamaño de muestra igual a **81** árboles.

6.4 RECORRIDOS PROPUESTOS PARA LA EVALUACIÓN DE EFICACIA Y PÉRDIDAS

Se propusieron dos métodos alternativos para seleccionar una muestra de árboles, que fuesen más prácticos en comparación a un muestreo aleatorio simple, con el fin de evaluar la jornada.

El primero de ellos consistió en escoger un árbol de la mitad del primer surco, y atravesar todos los surcos de ahí en adelante, siguiendo un recorrido en línea recta (Figura 1). La escogencia del primer árbol, se realizó con la ayuda del mapa de campo. El número de árboles establecido a evaluar en este recorrido fue el mismo que en el muestreo aleatorio simple, **81** árboles. Debido a que en las jornadas el número de surcos recolectados siempre fue mayor a 81, el criterio para la escogencia de los árboles fue el siguiente. De los primeros t surcos, se dejó de evaluar 1 árbol, donde t fue el entero más cercano a la razón $\frac{N}{N-81}$, siendo N el número de surcos recolectados en la jornada. Es decir, se evaluó un árbol en cada uno de los primeros $t-1$ surcos, en el siguiente surco no se evaluó ningún árbol, y así sucesivamente. Por esta razón, el número de árboles evaluados no fue exactamente 81 en cada jornada.

Figura 1. Recorrido en línea recta.

El segundo recorrido consistió en atravesar todos los surcos, formando una diagonal, empezando desde el primer árbol del primer surco y recorriendo en línea hasta el último árbol del último surco de la parcela (Figura 2). El número de árboles evaluado fue determinado con el mismo criterio presentado para el recorrido en recta.

Figura 2. Recorrido en diagonal.

6.5 RECOLECCIÓN DE INFORMACIÓN

Las labores de campo fueron realizadas en dos localidades del departamento de Caldas:

- Finca La Soledad, ubicada en el municipio de Palestina, Caldas. Latitud: 5° 02‘ Norte, Longitud: 75° 37‘ Oeste.

• Estación Central Naranjal, ubicada en el municipio de Chinchiná, Caldas. Latitud: 04° 58' Norte, Longitud: 75° 39' Oeste.

La caracterización de los lotes donde se evaluaron las jornadas de recolección se presentan en la tabla 8. Fueron lotes tecnificados (edad menor a 5 años) y la densidad de siembra entre los lotes es variable.

Tabla 8. Caracterización de los lotes donde se realizaron las evaluaciones de la recolección.

Sitio	Lote	Año de siembra	Densidad
La Soledad	3ª	2003	6.666 plantas/ha (1,5 x 1 mt)
	5B	2003 (50% zoca)	10.000 plantas/ha (2 x 1 mt)
	8	2001	7.400 plantas/ha (1,5 x 0,9 mt)
Naranjal	Castillo 3	2001	10.000 plantas /ha (1 x 1 mt)

La forma de contratación en la finca La Soledad fue al día (retribución económica establecida con anterioridad a la jornada), mientras que en la Estación Central Naranjal se pagó al contrato (pago por kilos recolectados).

Las 4 primeras jornadas se evaluaron en la finca La Soledad, en las siguientes fechas: 28 de marzo de 2005, lote 5B; 5 de abril de 2005, lotes 3ª y 5B; 6 de abril de 2005, lote 5B; y 13 de abril de 2005, lote 8. Las jornadas 5, 6, 7 y 8 se realizaron en la estación central Naranjal, los días 2, 13,16, 31 de mayo de 2005, respectivamente, en algunas parcelas de este lote.

En el período en que se realizaron las evaluaciones (marzo – junio), la participación en el total producido en el año en la zona central cafetera, es del 19%, según Alvarado y Moreno (1999).

Para la selección de los árboles correspondientes al muestreo aleatorio simple, se realizó el siguiente procedimiento:

- El día anterior a la evaluación se establecieron cuales y cuantos árboles corresponderían a la población objetivo, es decir se elaboró el marco de muestreo (mapa). Los árboles se numeraron de 1 a *N*, donde *N* fue el número de árboles recolectados en la jornada.

- Se realizó una selección aleatoria sin reemplazo de 81 números de 1 a *N*, con el fin de identificar los árboles que se debían evaluar.

- En los mapas correspondientes al marco de muestreo, se ubicaron los árboles seleccionados. Se realizó un plan para encontrarlos en el terreno, identificando tanto el surco donde se encontraba el árbol, como los árboles que había que recorrer desde el primer árbol del surco para llegar a él.

La evaluación en los árboles seleccionados mediante los recorridos alternativos propuestos, se realizó el mismo día de la ejecución del muestreo aleatorio simple, tomando en cuenta la consideración del número de surcos recolectados.

En cada jornada evaluada, para cada surco, se identificó al recolector que realizó su labor en sus árboles, con el fin de asociar un árbol evaluado a un recolector específico.

En cada uno de las técnicas de muestreo propuestas, se contó en cada árbol seleccionado, el número de frutos maduros, sobremaduros y secos en sus ramas, y el número de frutos en el suelo. Al finalizar cada una de estas técnicas de muestreo, se registró el tiempo empleado en su aplicación en campo.

6.6 ANÁLISIS DE LA INFORMACIÓN

6.6.1 Muestreo aleatorio simple.

- Estadística descriptiva.

- Estimación del error de muestreo alcanzado en cada jornada. Se obtiene despejando *E* de la fórmula propuesta para el error de muestreo para el promedio en la tabla 6.

$$E = 2s\sqrt{\frac{N-n}{n(N-1)}}$$

s es la desviación estándar estimada de la variable de interés en la jornada; *N* es el número de árboles recolectados; y *n* es el número de árboles seleccionados para la muestra.

6.6.2 Recorridos propuestos.

- Estadística descriptiva.

• Comparación de los tiempos aplicados por trayectoria y muestreo aleatorio simple. Para cada trayectoria, se realizó una prueba de hipótesis donde H_0 : $R = \frac{\overline{Y}}{\overline{X}} = 0{,}5$ y $H_1 : R \neq 0{,}5$, donde $\overline{Y}$ es el tiempo promedio a través de las jornadas al aplicar el recorrido, y $\overline{X}$ es el tiempo promedio al realizar un muestreo aleatorio simple. Se desea probar que la reducción en el tiempo de aplicación de un recorrido comparado con un muestreo aleatorio simple es aproximadamente del 50%.

• Pruebas de aleatorización. Para cada indicador se realizaron dos pruebas de aleatorización, cuyas hipótesis nulas fueron "Los valores obtenidos en la evaluación no dependen del tipo de muestreo aplicado". El objetivo es determinar si existe evidencia estadística de que el método de muestreo con el que se evalúe el proceso de recolección incide o no, en los valores obtenidos en la muestra.

El análisis fue realizado para cada sitio donde se realizaron las jornadas de recolección, tomando en cuenta las diferencias del proceso entre los dos sitios.

En la primera prueba, los grupos de interés fueron los datos obtenidos por muestreo aleatorio simple, y los obtenidos en el recorrido recta. En la segunda prueba, se comparó el muestreo aleatorio simple y el recorrido diagonal. Se comparan los valores de los recorridos propuestos con los del muestreo aleatorio simple, debido que este es un muestreo probabilístico cuya confiabilidad es conocida por la aleatoriedad de sus muestras.

El estadístico de prueba es la suma de cuadrados entre grupos presentada en el marco teórico para un análisis de varianza con un factor de clasificación. El factor de clasificación es la jornada de evaluación. El proceso de aleatorización se realizó dentro de los valores que tomaba este factor de clasificación. El número de

reordenamientos aleatorios dentro de cada valor de la variable jornada, fue de 5000, para ambas pruebas, atendiendo la recomendación de Howell (2002).

Mediante los reordenamientos aleatorios, se construyeron las distribuciones de referencia del estadístico de prueba. Se estimó el respectivo valor *p*, y se definió como nivel de significancia $\alpha = 0{,}05$, es decir que se rechazará la hipótesis nula si el porcentaje de reordenamientos aleatorios que tienen un estadístico de prueba mayor al obtenido con los datos originales, es menor al 5%.

El algoritmo realizado en SAS tuvo como base el presentado por David Cassel para pruebas de aleatorización[17]. Tal algoritmo se presenta como en el anexo A, al final del documento.

[17] CASSELL, D. L. A Randomization-test Wrapper for SAS® PROCs. [Online]. http://www2.sas.com/proceedings/sugi27/p251-27.pdf. (Consultado en Marzo de 2006).

7. RESULTADOS Y DISCUSIÓN

En resumen, el marco muestral y el número de operarios por jornada se presenta en la tabla 9. En La Soledad el número de árboles por jornada fue menor comparado con Naranjal, al igual que el número de operarios. Los mapas de los lotes donde se realizó la recolección en las jornadas se presentan en el anexo B, al final del documento.

Tabla 9. Número de árboles recolectados (población *N*) y número de operarios que realizaron la labor por jornada.

Sitio	Jornada	Número de operarios	*N*
La Soledad	1	8	4097
	2	7	4138
	3	7	2549
	4	6	4003
Naranjal	1	15	8414
	2	15	10302
	3	15	8414
	4	15	8059

Se presentan los resultados del muestreo aleatorio simple, respecto a las estimaciones de los parámetros y los errores de estimación obtenidos en las jornadas. Luego se analizan los resultados de los recorridos propuestos a través de las jornadas.

7.1 MUESTREO ALEATORIO SIMPLE

7.1.1 Estadística descriptiva.

Eficacia. Los promedios del número de frutos maduros dejados en el árbol fueron estadísticamente menores a cinco (estándar establecido por Bustillo, 2002, para manejo integrado de la broca), excepto en la jornada 3 en Naranjal, para la cual el intervalo de confianza del promedio contiene al valor 5. En este sitio las varianzas estimadas para cada una de estas variables fueron mayores a las presentadas en La Soledad.

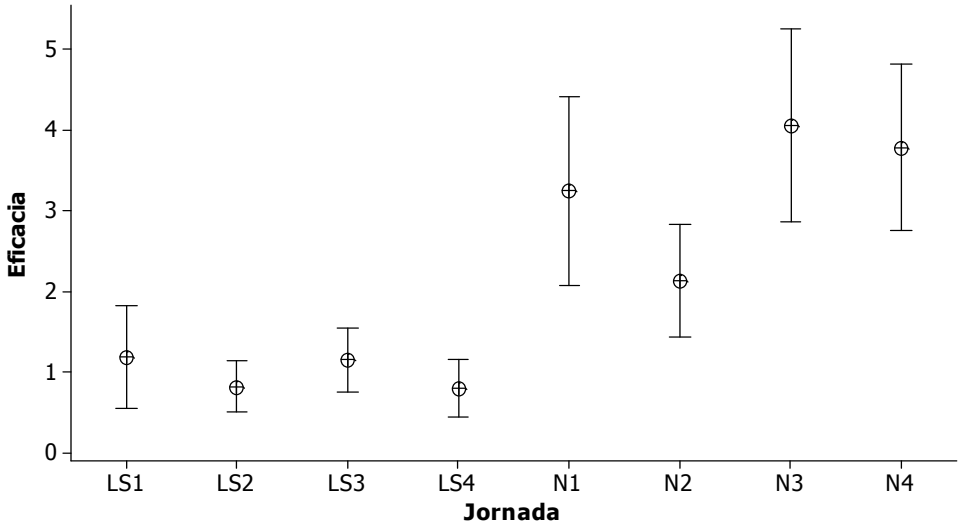
Tabla 10. Promedios y límites de confianza (α = 0,95), para el número de frutos dejados en el árbol (eficacia), por jornada.

Sitio	Jornada	EFICACIA		
		Promedio	Varianza	Intervalo de confianza
La Soledad	1	1,2	9,5	0,6 – 1,8
	2	0,8	2,2	0,5 – 1,1
	3	1,2	3,3	0,8 – 1,5
	4	0,8	2,6	0,4 – 1,2
Naranjal	1	3,2	30,1	2,1 – 4,4
	2	2,1	10,3	1,4 – 2,8
	3	4,1	30,4	2,9 – 5,3
	4	3,8	21,1	2,8 – 4,8

En la figura 3, se observa que las jornadas en Naranjal tienen promedios de frutos maduros dejados en el árbol mayores a los de La Soledad, (intervalo de confianza

de 95%). Como se mencionó anteriormente, se observan intervalos de confianza para el promedio de mayor amplitud, en Naranjal.

Figura 3. Diagrama de intervalos para el promedio del número de frutos maduros sin recolectar (eficacia), por jornada.

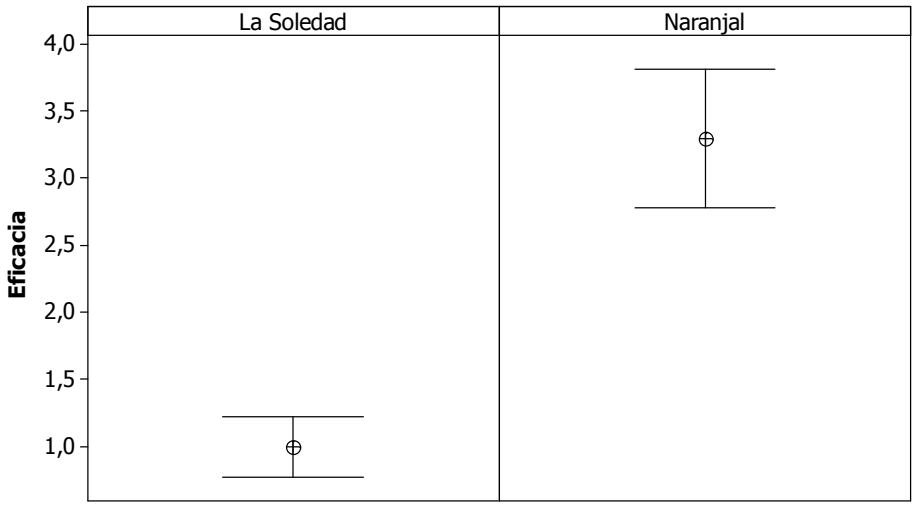


El indicador eficacia fue estadísticamente mayor en Naranjal que en la Soledad (Tabla 11, Figura 4), lo que reafirma la conveniencia de realizar un análisis separado por sitio respecto a este indicador.

Tabla 11. Promedios y límites de confianza (α = 0,95), para el número de frutos dejados en el árbol (eficacia), por sitio.

Sitio	**EFICACIA**		
	Promedio	**Varianza**	**Intervalo de confianza**
La Soledad	1,0	4,5	0,8 – 1,2
Naranjal	3,3	23,4	2,8 – 3,8

Figura 4. Diagrama de intervalos para el promedio del número de frutos maduros sin recolectar (eficacia), por sitio.



Las distribuciones del número de frutos maduros fueron asimétricas positivamente (coeficientes de asimetría mayores a 0), y en cada una se rechaza la hipótesis de que sus datos se ajustan a una distribución normal (tabla 12).

Pérdidas. En Naranjal se presentaron promedios mayores que en La Soledad, y en tres de las cuatro jornadas estos fueron estadísticamente mayores a cinco. En La Soledad solo en una jornada el promedio fue mayor a cinco frutos por árbol. En La Soledad, los recolectores tuvieron a su disposición una bolsa denominada "líchigo", donde podían almacenar los frutos verdes que hubiesen recogido en su tarea. Esta circunstancia explica la diferencia entre los sitios, en cuanto la variable número de frutos dejados en el suelo (Tabla 13).

Tabla 12. Coeficiente de asimetría y prueba de normalidad Anderson Darling para el número de frutos maduros dejados en el árbol (eficacia).

Sitio	Jornada	EFICACIA	
		Coef. de Asimetría	Anderson Darling (Valor *p*)
La Soledad	1	5,0	< 0,005
	2	2,5	< 0,005
	3	1,8	< 0,005
	4	3,3	< 0,005
Naranjal	1	3,2	< 0,005
	2	2,7	< 0,005
	3	2,7	< 0,005
	4	2,1	< 0,005

Tabla 13. Promedios y límites de confianza (α = 0,95), para el número de frutos dejados en el suelo (pérdidas), por jornada.

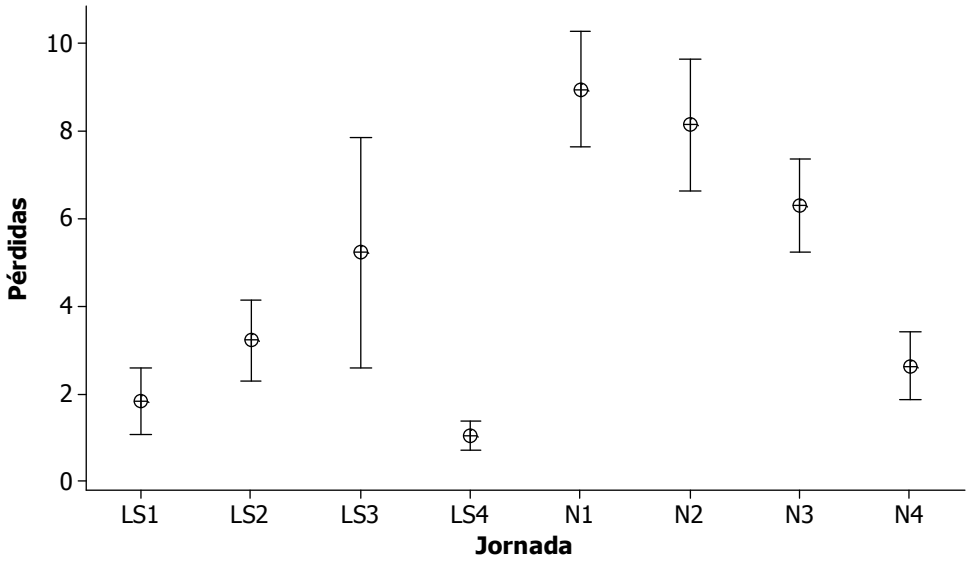
Sitio	Jornada	PERDIDAS		
		Promedio	Varianza	Intervalo de confianza
La Soledad	1	1,8	13,6	1,1 – 2,6
	2	3,2	18,6	2,3 – 4,1
	3	5,2	149,1	2,6 – 7,8
	4	1,0	2,2	0,7 – 1,4
Naranjal	1	8,9	38,3	7,6 – 10,3
	2	8,1	47,1	6,6 – 9,6
	3	6,3	23,9	5,2 – 7,3
	4	2,6	11,9	1,9 – 3,4

En cinco de las 8 jornadas, la varianza fue mayor en comparación la obtenida para el indicador Eficacia en esa misma jornada (Tablas 11 y 13). La mayor varianza se

obtuvo en la jornada 3 en La Soledad, donde en un árbol seleccionado se encontraron 98 frutos en el suelo. Se puede observar, que en esta jornada, se presenta al intervalo de confianza de mayor longitud para el promedio (Figura 5).

Los intervalos de confianza para los promedios, muestran que estos son diferentes entre algunas de las jornadas (Figura 5).

Figura 5. Diagrama de intervalos para el promedio del número de frutos dejados en el suelo (pérdidas), por jornada.

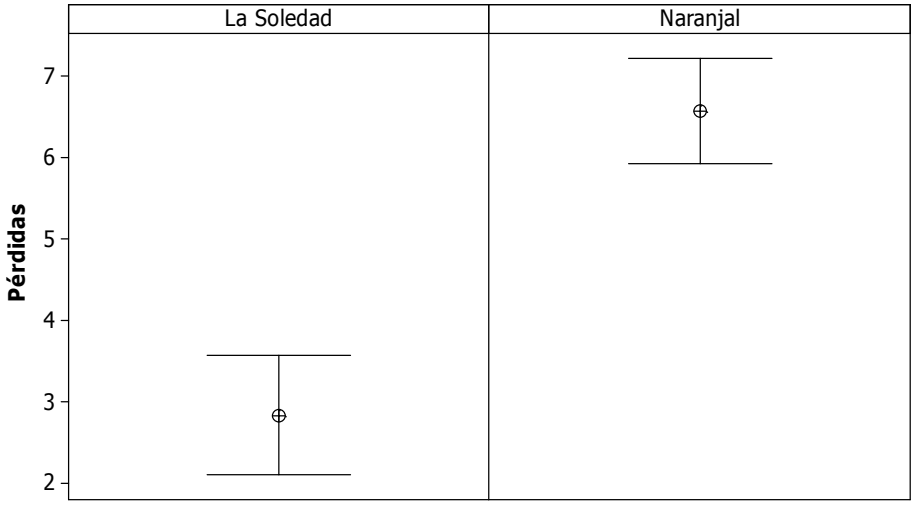


Los sitios son diferentes estadísticamente respecto al promedio del número de frutos en el suelo por árbol (Tabla 14, Figura 6).

Tabla 14. Promedios y límites de confianza (α = 0,95), para el número de frutos dejados en el suelo (pérdidas), por sitio.

Sitio	PERDIDAS		
	Promedio	**Varianza**	**Intervalo de confianza**
La Soledad	2,8	47,8	2,1 – 3,6
Naranjal	6,6	36,1	5,9 – 7,2

Figura 6. Diagrama de intervalos para el promedio del número de frutos dejados en el suelo (pérdidas), por sitio.



Las distribuciones de los datos en cada jornada, son asimétricos positivamente, y tienen una distribución diferente a la normal (Tabla 15).

Tabla 15. Coeficiente de asimetría y prueba de normalidad Anderson Darling para el número de frutos maduros dejados en el suelo (pérdidas).

Sitio	Jornada	EFICACIA	
		Coef. de Asimetría	Anderson Darling (Valor *p*)
La Soledad	1	5,7	< 0,005
	2	1,7	< 0,005
	3	5,8	< 0,005
	4	1,7	< 0,005
Naranjal	1	0,6	< 0,005
	2	2,0	< 0,005
	3	1,1	< 0,005
	4	3,0	< 0,005

7.1.2 Verificación del error de estimación. En todas las jornadas, el error de estimación para el promedio de frutos maduros dejados por árbol fue menor a 2, que fue el valor establecido (Tabla 16).

En 7 de las 8 evaluaciones, el límite para el error de estimación del promedio de frutos dejados en el suelo por árbol, fue menor a 2 (Tabla 17). Solo en la jornada 3 en La Soledad el error obtenido fue mayor a este valor, es decir, el porcentaje de evaluaciones para el cual se cumplió el error de estimación establecido fue del 87,5%, para la variable frutos dejados en el suelo y del 100%, para frutos dejados en el árbol.

Tabla 16. Límite para el error de estimación del promedio del número de frutos maduros dejados en el árbol (eficacia).

Sitio	Jornada	Varianza	*N*	Error de estimación
La Soledad	1	9,5	4097	0,6
	2	2,2	4138	0,3
	3	3,3	2549	0,4
	4	2,6	4003	0,4
Naranjal	1	30,1	8414	1,2
	2	10,3	10302	0,7
	3	30,4	8414	1,2
	4	21,1	8059	1,0

Tabla 17. Límite para el error de estimación del promedio del número de frutos dejados en el suelo (pérdidas).

Sitio	Jornada	Varianza	*N*	Error de estimación
La Soledad	1	13,6	4097	0,8
	2	18,6	4138	0,9
	3	149,1	2549	2,6
	4	2,2	4003	0,3
Naranjal	1	38,3	8414	1,3
	2	47,1	10302	1,5
	3	23,9	8414	1,1
	4	11,9	8059	0,8

7.2 RECORRIDOS PROPUESTOS

7.2.1 Estadística descriptiva. En las tablas 18 y 19 se presentan los promedios y las varianzas por jornada para los indicadores eficacia y pérdidas respectivamente, para los recorridos propuestos. Se puede observar como el valor de la varianza es muy variable a través de las jornadas, en especial, en el caso del indicador pérdidas.

Tabla 18. Promedio y varianza del número de frutos dejados en el árbol (eficacia), para cada recorrido.

Sitio	Jornada	Recta		Diagonal	
		Promedio	Varianza	Promedio	Varianza
La Soledad	1	0,8	3,8	0,6	1,7
	2	1,8	9,4	1,6	7,9
	3	1,2	4,1	2,4	57,1
	4	1,5	9,5	1,6	12,7
Naranjal	1	3,2	16,9	2,9	15,0
	2	3,4	13,6	3,5	24,0
	3	3,9	15,1	5,1	54,6
	4	5,0	35,2	6,0	43,5

Tabla 19. Promedio y varianza del número de frutos dejados en el suelo (pérdidas), para cada recorrido.

Sitio	Jornada	Recta		Diagonal	
		Promedio	Varianza	Promedio	Varianza
La Soledad	1	1,0	2,1	1,9	12,2
	2	6,2	88,8	5,2	88,6
	3	3,4	52,5	3,7	27,2
	4	1,1	3,1	0,7	0,8
Naranjal	1	10,8	42,4	11,9	63,6
	2	10,8	92,6	7,1	41,4
	3	9,5	47,5	7,5	40,3
	4	4,1	16,8	4,6	34,3

7.2.2 Tiempos de ejecución. Respecto a los tiempos de aplicación, estos se reducen a casi la mitad realizando los recorridos propuestos en comparación al muestreo aleatorio simple (Tabla 20).

Tabla 20. Tiempos de aplicación de los diferentes tipos de evaluación.

		Método		
Sitio	**Jornada**	**Aleatorio**	**Recta**	**Diagonal**
La Soledad	1	2 h 23 m	42 m	46 m
	2	3 h 10 m	1 h 30 m	1 h 23 m
	3	2 h 13 m	1 h 08 m	1 h 29 m
	4	2 h 25 m	1 h 16 m	1 h 20 m
Naranjal	1	2 h 23 m	1 h 19 m	1 h 32 m
	2	2 h 38m	1 h 45 m	1 h 28 m
	3	2 h 58 m	1 h 40 m	1 h 27 m
	4	2 h 21 m	1 h 36 m	1 h 50 m

Se estimaron las razones del tiempo empleado en realizar cada recorrido y el tiempo de realizar muestreo aleatorio simple (Tabla 21). También se estimaron los intervalos de confianza. Al aplicar un recorrido se reduce el 50% del tiempo en la aplicación en comparación al muestreo aleatorio simple. Esto sin tener en cuenta el tiempo que demanda la planeación del muestreo aleatorio simple con anterioridad a una jornada.

Tabla 21. Razones de los tiempos empleados al utilizar un recorrido en comparación a un muestreo aleatorio simple.

Método	*R**	Error estándar	Límite inferior	Límite superior
Recta	0,53	0,04	0,43	0,63
Diagonal	0,55	0,05	0,43	0,66

***** Razón: tiempo recorrido / tiempo muestreo aleatorio simple

Se cumplió entonces el objetivo de la disminución del tiempo en la aplicación de los recorridos propuestos en campo. Adicionalmente, la planeación del muestreo aleatorio simple con anterioridad a una jornada, demandó un tiempo que no se tuvo en cuenta al momento de la comparación.

Se realizó una prueba de diferencia de medias entre los tiempos de ambos recorridos, para establecer si alguno de los dos requiere menos tiempo en su ejecución. No existe evidencia para rechazar la hipótesis de igualdad en los tiempos de ejecución de los recorridos propuestos (Tabla 22).

Tabla 22. Significancia de la hipótesis H_0: $\mu_{Recta} = \mu_{Diagonal}$ vs. H_1: $\mu_{Recta} \neq \mu_{Diagonal}$, para los tiempos de cada recorrido.

	Media (horas)	Error estándar	*t*	Valor *p*
Diferencia en los tiempos (Recta – Diagonal)	- 0,04	0,08	-0,48	0,644

7.2.3 Pruebas de aleatorización. El interés al realizar las pruebas de aleatorización es verificar si los valores obtenidos en un recorrido determinado son equivalentes a los de un muestreo aleatorio simple, para los indicadores eficacia y pérdidas.

A continuación se presenta el análisis para cada uno de los sitios.

- **La Soledad.** Los valores *p* alcanzados mediante la comparación de los valores del recorrido recta y el muestreo aleatorio simple de ambos indicadores, no fueron significativos para rechazar la hipótesis de no diferencia entre ellos (Tabla 23), considerando $\alpha = 0,05$.

Para el recorrido en diagonal, se rechaza la hipótesis nula de no diferencia entre los valores de este recorrido y el muestreo aleatorio simple para el indicador Eficacia, más no se rechaza la hipótesis nula en el caso del indicador Pérdidas (Tabla 23).

Tabla 23. Valores *p* de las pruebas de aleatorización para cada recorrido en La Soledad.

Recorrido	Indicador	Estadístico	Valor *p*
Recta	Eficacia	20,99	0,05
	Pérdidas	0,12	0,96
Diagonal	Eficacia	51,14	0,04
	Pérdidas	0,00	1,00

Las distribuciones de los estadísticos de prueba para La Soledad se ilustran en las figuras 7, 8, 9 y 10. El valor del estadístico de prueba obtenido es encerrado en un círculo rojo.

Figura 7. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido recta, indicador Eficacia, en La Soledad.

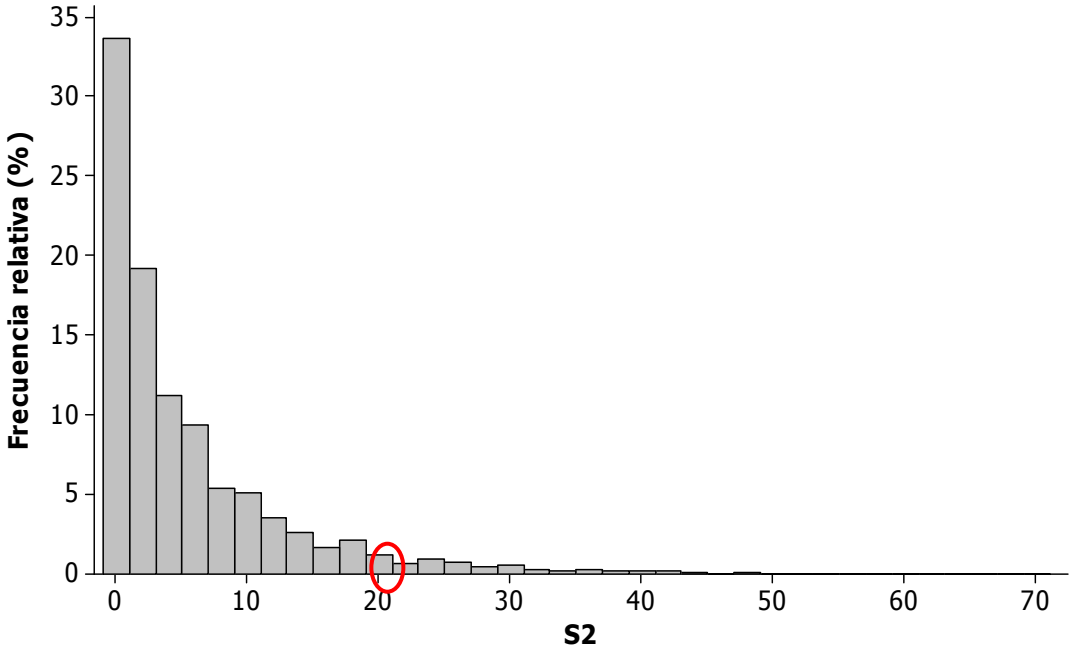


Figura 8. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido recta, indicador Pérdidas, en La Soledad.

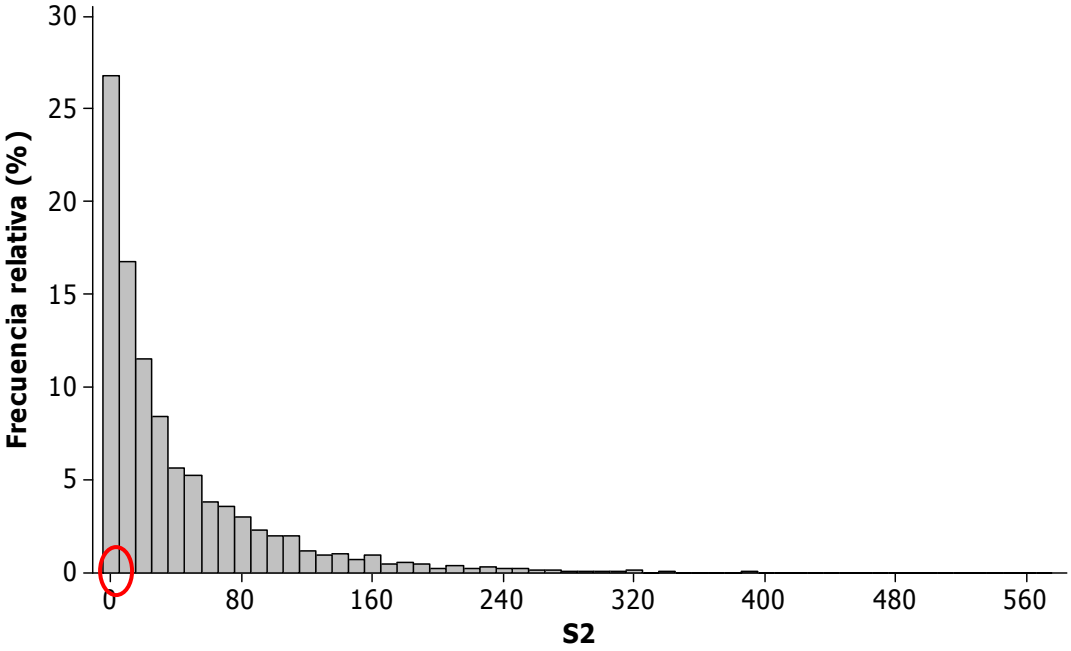


Figura 9. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido diagonal, indicador Eficacia, en La Soledad.

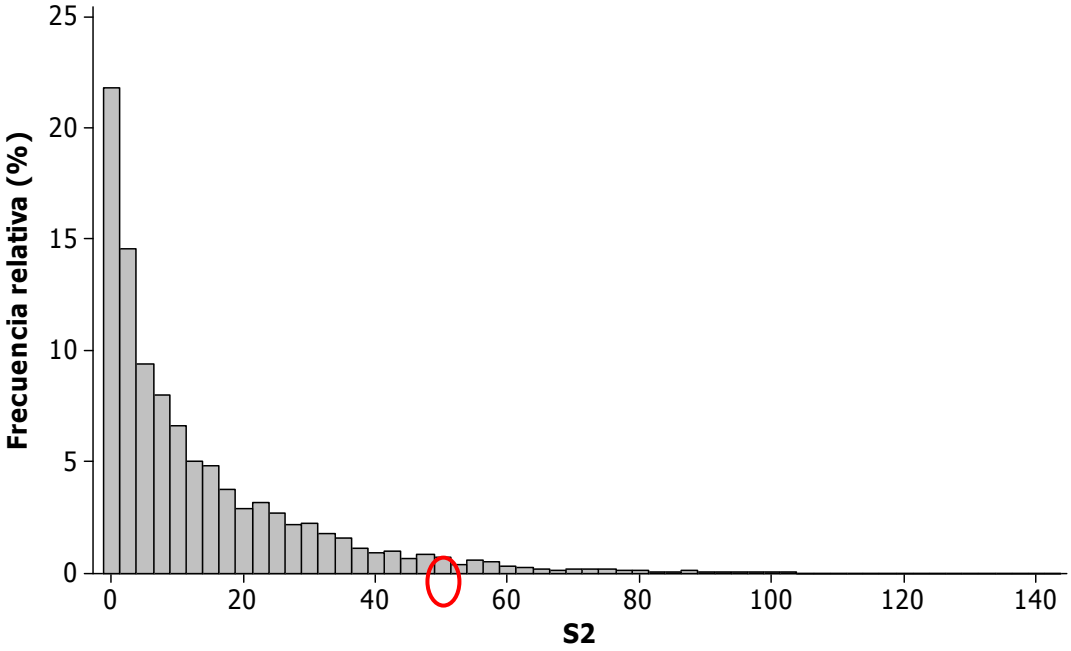
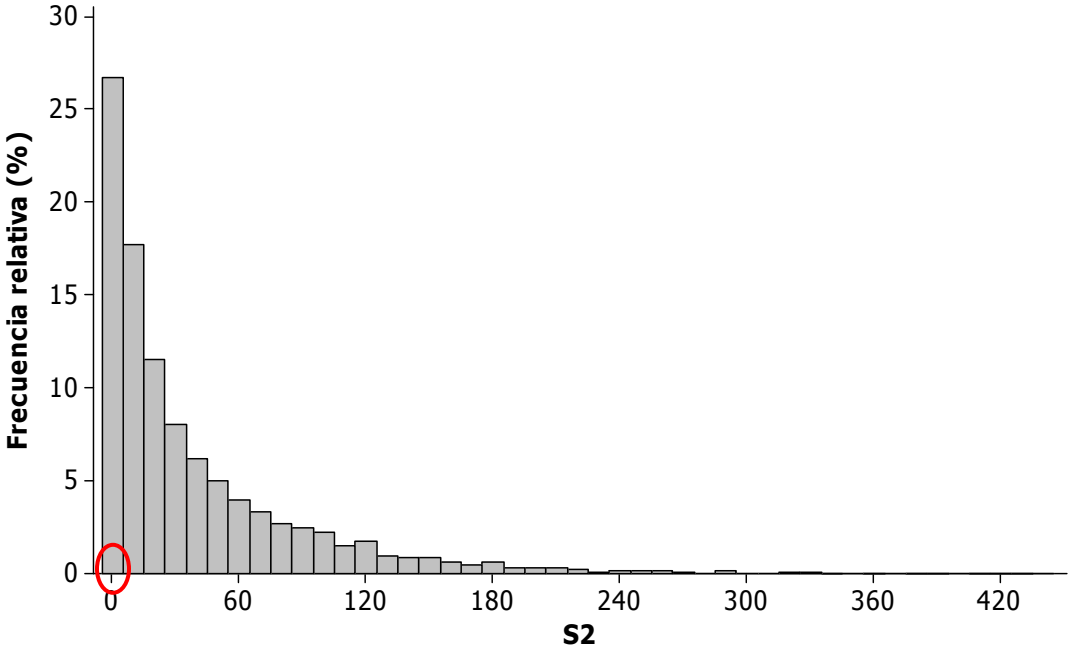


Figura 10. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido diagonal, indicador Pérdidas, en La Soledad.



- **Naranjal.** En la comparación de los valores obtenidos por el recorrido en Recta y el muestreo aleatorio simple, no se rechaza la hipótesis de no diferencia entre ellos, para el caso del indicador Eficacia. Sin embargo el valor *p* alcanzado para el indicador Pérdidas es 0, es decir el estadístico es significativo (Tabla 24).

Para ambos indicadores, los estadísticos de prueba de la comparación entre el recorrido en Diagonal y el muestreo aleatorio simple, son significativos, es decir se rechaza la hipótesis de no diferencia entre los valores de ambos métodos (Tabla 24).

Tabla 24. Valores *p* de las pruebas de aleatorización para cada recorrido en Naranjal.

Recorrido	Indicador	Estadístico	Valor *p*
Recta	Eficacia	55,39	0,11
	Pérdidas	921,20	0,00
Diagonal	Eficacia	210,29	0,01
	Pérdidas	267,11	0,01

Las distribuciones de los estadísticos de prueba para Naranjal se ilustran en las figuras 11, 12, 13 y 14.

Figura 11. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido recta, indicador Eficacia, en Naranjal.

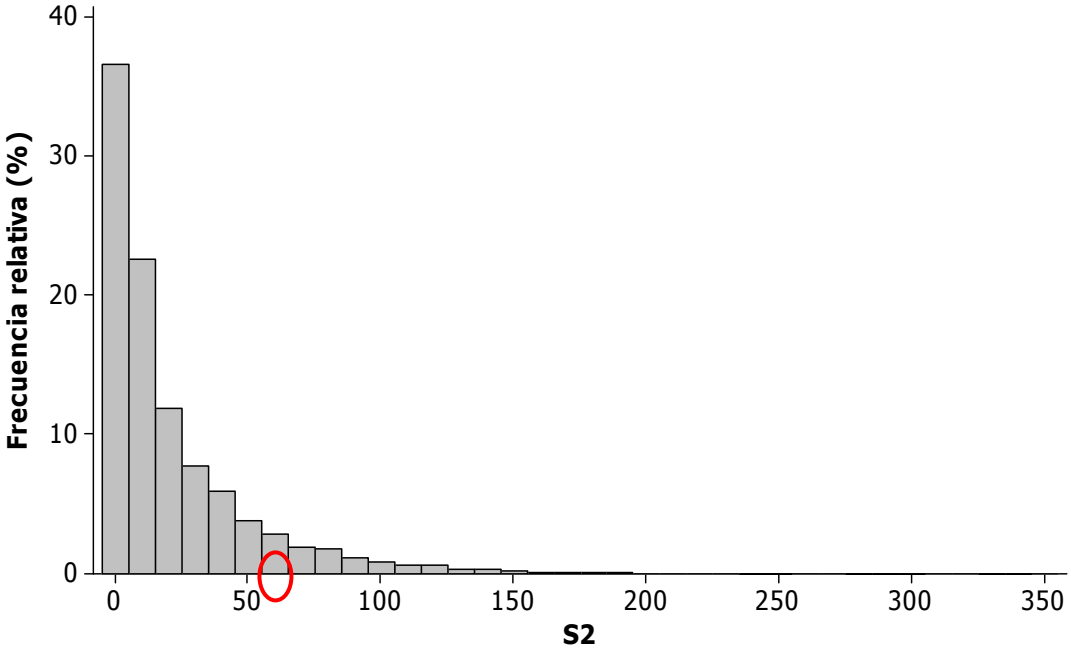


Figura 12. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido recta, indicador Pérdidas, en Naranjal.

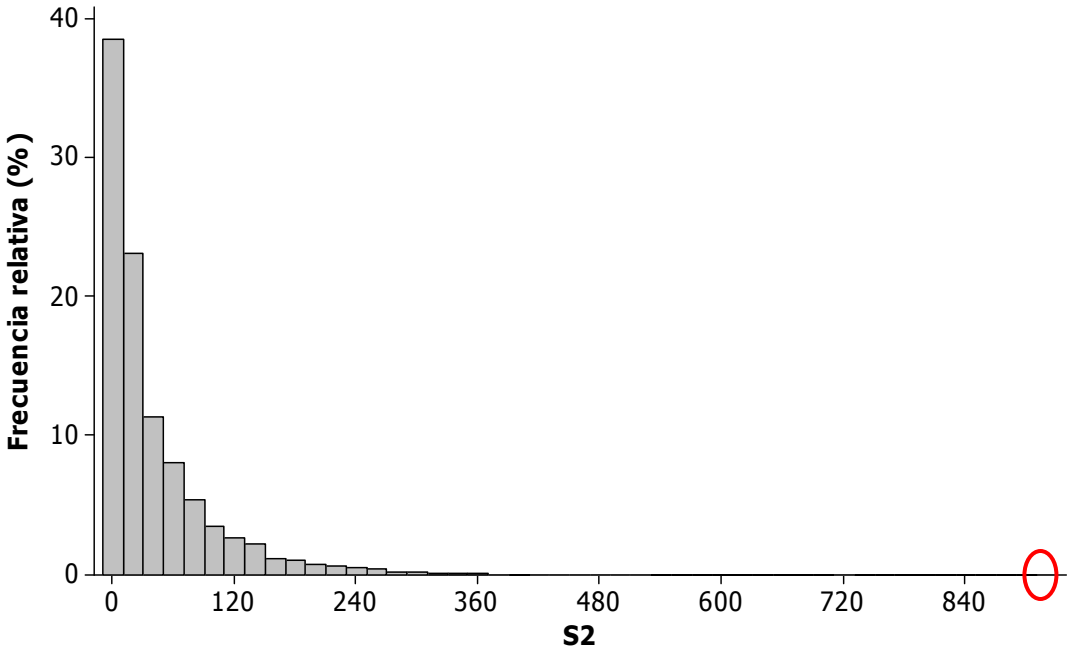


Figura 13. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido diagonal, indicador Eficacia, en Naranjal.

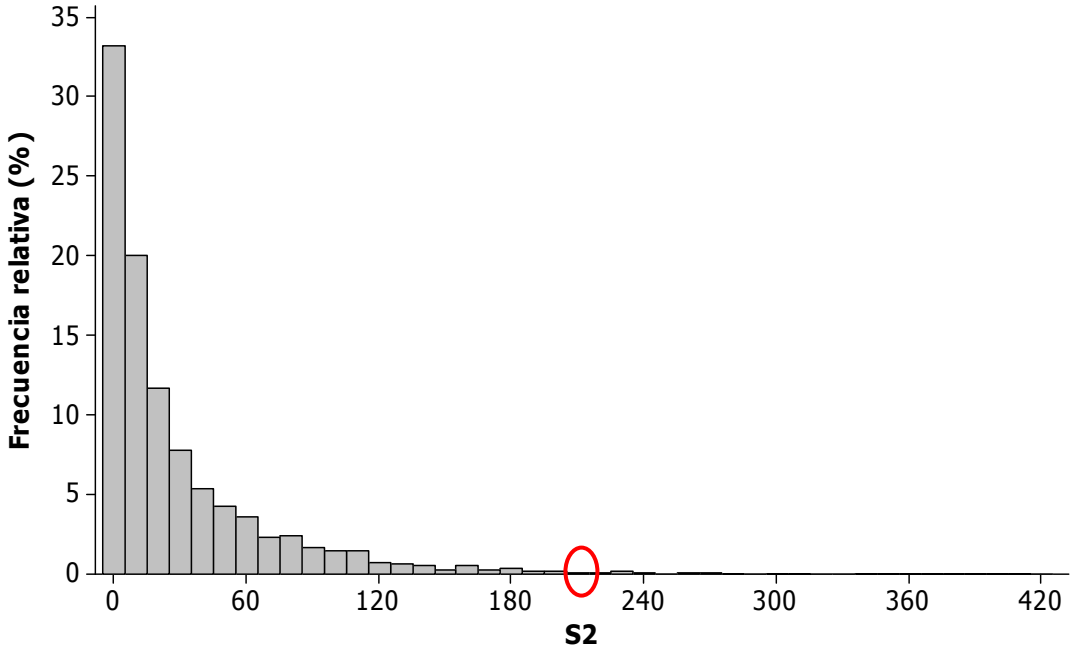
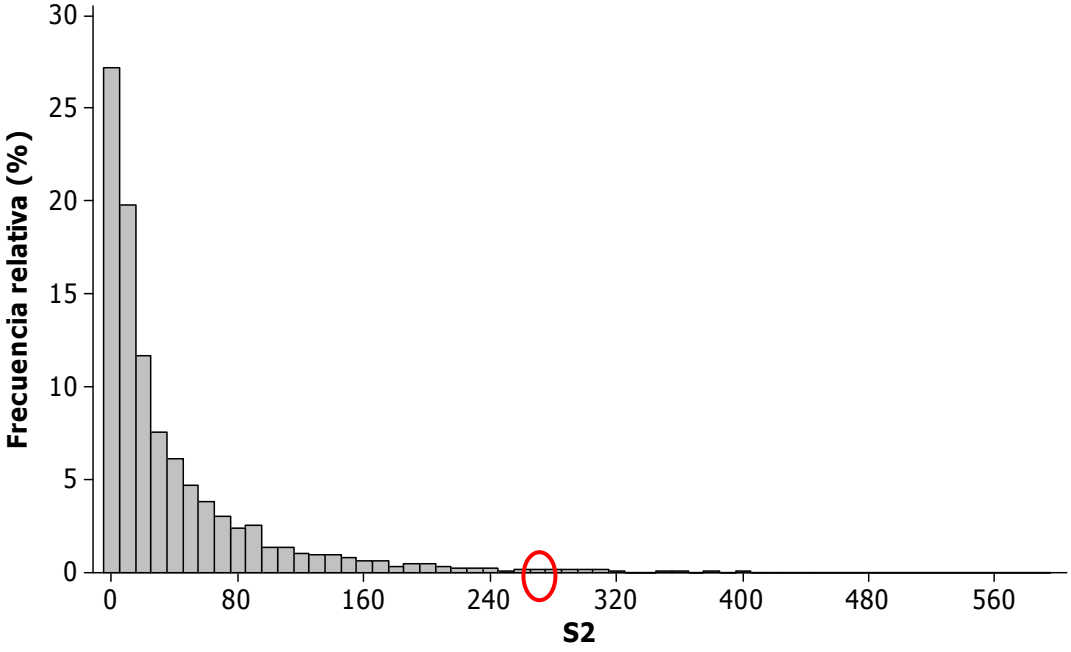


Figura 14. Distribución del estadístico de prueba para la comparación del muestreo aleatorio simple y el recorrido diagonal, indicador Pérdidas, en Naranjal.



En conclusión, para α = 0,05, no se rechaza la hipótesis nula de que los valores obtenidos por el recorrido recta y el muestreo aleatorio simple respecto al indicador Eficacia no son diferentes, en ambos sitios. En Naranjal se rechaza esta hipótesis para el indicador Pérdidas.

La comparación de los valores del recorrido diagonal y muestreo aleatorio simple para cada indicador por sitio, lleva a la conclusión de que se rechaza la hipótesis de no diferencia entre los valores de ambos métodos, con la excepción del caso del indicador Pérdidas en La Soledad.

8. CONCLUSIONES

- La ejecución de un muestreo aleatorio simple en un cafetal no es práctica. Exige la disponibilidad de un mapa detallado donde se identifique cada árbol en el terreno, además se debe realizar la selección aleatoria de las unidades muestrales, y luego la ubicación de ellas en campo. Este proceso se debe realizar para cada jornada de recolección, siendo no práctico, tomando en cuenta que se debe disponer de los resultados de la evaluación de manera rápida para tomar decisiones oportunas.

- Es importante destacar la idea de reducir el número de frutos sin recolectar. Es una inversión que se ha realizado, además de que se disminuyen los costos del manejo integrado de broca.

- El error de estimación establecido se cumplió con el tamaño de muestra propuesto (81 árboles), en el 100 % de las jornadas para el indicador eficacia, y en el 85 % de las mismas para el indicador pérdidas.

- Con base en los resultados de las pruebas de aleatorización para cada caso, se concluye que no existen diferencias entre las observaciones obtenidas por el muestreo aleatorio simple y el recorrido en recta, para la evaluación del número de frutos dejados en el árbol, en los dos sitios; se rechaza la hipótesis para el indicador Pérdidas para Naranjal. Para los valores obtenidos por el recorrido en

diagonal, solamente para el indicador Pérdidas en La Soledad no se rechaza la hipótesis nula; para eficacia en ambos sitios, y pérdidas en Naranjal, se encontraron diferencias significativas con los valores del muestreo aleatorio simple.

- Para ambos recorridos, se cumple el objetivo de reducir los tiempos de aplicación con respecto al tiempo utilizado para aplicar muestreo aleatorio simple (aproximadamente un 50%).

- Las pruebas de aleatorización proponen un acercamiento a los problemas diferente al ofrecido por las pruebas paramétricas. Esto permite realizar pruebas sobre datos que no cumplen algunos supuestos de los exigidos por las pruebas paramétricas, pero también restringe las conclusiones. La discusión está abierta sobre la bondad de estas pruebas y este tema debería ser más difundido en el ambiente universitario, como propone Lunneborg[18].

- El tema de la recolección es difícil y complejo, porque involucra agentes humanos, agronómicos, socioeconómicos y culturales. Sin embargo, este proceso debe ser estudiado para poder establecer los factores que son determinantes en su mejoramiento, lo que llevará a una reducción de los costos relacionados con el proceso de recolección.

[18] A.P.S How Random is That? The observer. Sept. 2005. V. 18, No. 9. (Online) http://www.psychologicalscience.org/observer/getArticle.cfm?id=1838

9. RECOMENDACIONES

- El método de evaluación debe ser práctico. Sin embargo, también debe tener sustento científico. El número de árboles propuesto en el trabajo en la evaluación de una jornada por un muestreo aleatorio simple (90 árboles) parece exagerado para un caficultor, considerando que se debe realizar la transferencia de estos resultados. Se recomienda estudiar tamaños de muestra por parcela, o por tiempo trabajado, para realizar una evaluación que permita tomar decisiones a tiempo sobre el proceso.

- Los resultados de este estudio no son definitivos. Existe la posibilidad de proponer otros métodos de evaluación. En el estudio se encontró que un recorrido (recta) no tuvo resultados diferentes a la de un muestreo formal, en dos localidades determinadas. Se debe realizar la validación de los resultados, en otras localidades.

- Antes de realizar exigencias a los recolectores, se les debe ofrecer el ambiente adecuado para realizar sin mayores dificultades su trabajo, como por ejemplo, densidades de siembra, y árboles podados, que le permitan al recolector transitar sin dificultad por las calles del cafetal. Esto también permitirá un mejor control del proceso tanto en la facilidad para el desplazamiento en el lote, como en la evaluación visual de los indicadores.

BIBLIOGRAFÍA

ALVARADO A., G. y MORENO R., L. G. Cómo se distribuye anualmente la cosecha de las variedades Caturra y Colombia? En: Avances Técnicos Cenicafé (Colombia). No. 260 (1999); p. 1-4.

ARCILA P., J.; BUHR, L.; BLEIHOLDER, H.; HACK, H.; y WICKE, H. Aplicación de la escala BBCH ampliada para la descripción de las fases fenológicas del desarrollo de la planta de café *Coffea sp*. En: Boletín Técnico Cenicafé (Colombia). No. 23 (2001); p. 1-31.

ARENAS A., J. E.; ARIAS L., J. A.; BUSTAMANTE G., F. J.; DIAZ B., Y; MARIN A., H. F.; y RODRIGUEZ D., H. G. Evaluación de la calidad de la recolección del fruto del cafeto, sus implicaciones técnicas y socio-económicas en diez municipios del departamento de Caldas. Manizales, 1997, 45 p. (Estudio de caso). Universidad de Caldas, Programa Agronomía – Comité Departamental de Cafeteros de Caldas.

BEHAR G., R. y YEPES A., M. Estadística: un enfoque descriptivo. Cali: Universidad del Valle, División de Ingeniería, 1996. 235 p.

BUSTILLO P., A. E. El manejo de cafetales y su relación con el control de la broca del café en Colombia. En: Boletín Técnico Cenicafé (Colombia). No. 24, (2002); p. 1-40.

CASTAÑO D., B. y JARAMILLO H., O. Administración de la cosecha (s. p. e.).

CENTRO NACIONAL DE INVESTIGACIONES DE CAFÉ, Colombia. Qué hacer después de terminada la cosecha. En: Brocarta (Colombia). No. 27 (1995); p. 1-3.

_____. Doce maneras de mejorar los ingresos en las fincas cafeteras. En: Avances Técnicos Cenicafé (Colombia). No. 255 (1998); p. 1-7.

COCHRAN, W. G. Técnicas de muestreo. México: Continental, 1978. 507 p.

CONGEDO, M. The randomization theory: An Introduction [en línea]. [USA]: 2000 [citado en febrero 20 de 2006]. Disponible en Internet: <www.novatecheeg.com/Congedo/The Randomization Theory Marzo 2000.ppt>

CONOVER, W. Practical nonparametric statistics. 3 ed. USA: John Wiley & Sons, 1999. 584p.

CHAMORRO T., G. E.; CARDENAS M., R. y HERRERA H., A. Evaluación económica y de la calidad en taza del café proveniente de diferentes sistemas de recolección manual, utilizables como control en cafetales infestados de *Hypothenemus hampei*. En: Cenicafé (Colombia), 1995. Vol. 46, No. 3 (1995); p.164-175.

CHOU, Y-L. Análisis Estadístico. México: Interamericana, 1975. 808 p.

DIAZ B., Y. y MARIN A., H. F. Evaluación de los frutos dejados en la recolección durante un ciclo productivo del cultivo en dos municipios del Departamento de Caldas. Manizales, 1999, 93 p. Tesis (Ingeniero Agrónomo). Universidad de Caldas. Facultad de Ciencias Agropecuarias, 1999.

DIAZ-URIARTE, R. The analysis of cross-over trials in animal behavior experiments: review and guide to the statistical literature [en línea]. Departments of Zoology and Statistics University of Wisconsin-Madison. USA, 2001 [citado en 10 de marzo de 2006]. Disponible en Internet: <http://samizdat.mines.edu/diaz-uriarte/cross-over.pdf>

DUQUE O., H. Como reducir los costos de producción en la finca cafetera. Chinchiná : Cenicafé, 2002. 85 p.

_____. Estimación de la función de ingreso del recolector de café, al contrato. Seminario Julio 25, 2003. Chinchiná: Cenicafé, 2003.

EDGINGTON, E. S. Randomization tests. 3 ed. New York: Marcel Dekker, 1995. 409 p.

FEDERACIÓN NACIONAL DE CAFETEROS DE COLOMBIA. Costos de producción de café : Zona central cafetera. Santafé de Bogotá, Federacafé. Gerencia Técnica. División de producción y desarrollo social. Café y administración rural, Santafé de Bogotá, 2000. 13 p.

FRACICA N., G. Modelo de simulación en muestreo. Colombia: Universidad de La Sabana, 1998. 128 p. (Serie Investigación – Docencia; no. 18)

HERNANDEZ S., R.; FERNANDEZ C., C.; y BAPTISTA L., P. Metodología de la investigación. 2 ed. México: McGraw-Hill, 1998. 501p.

HINCAPIE N., M. Diagnóstico de la administración de la cosecha de café en fincas empresariales de la zona central cafetera de Colombia. Manizales, 2005, 104 p. Tesis (Ingeniero Agrónomo). Universidad de Caldas. Facultad de Ciencias Agropecuarias, 2005.

HOWELL, D. C. Resampling Statistics: Randomization and the Bootstrap [en línea]. University of Vermont. USA, 2002 [citado en febrero 13 de 2006]. Disponible en Internet: <http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html>

KISH, L. Muestreo de encuestas. México: Trillas, 1979. 739 p.

KLÍNGER, R. La estadística en las investigaciones. Cali: Universidad del Valle, 2003. (Notas de clase).

LATORRE E., E. Teoría general de sistemas: aplicada a la solución integral de problemas. Cali (Colombia): Universidad del Valle, 1996. 217 p.

MARTINEZ R., R. A. Estudio de la cosecha manual del café en condiciones de altas pendientes. Cali, 2004, 84 p. Tesis (Ingeniero Agrícola). Universidad del Valle. Facultad de Ingeniería.

MÉNDEZ A., C. E. Metodología: diseño y desarrollo del proceso de investigación. Bogotá: McGraw-Hill, 2001. 247 p.

MONTOYA R., E. C. Disminución de los costos de recolección mediante la cosecha manual asistida – Enfoque teórico. En: CENTRO NACIONAL DE INVESTIGACIONES DE CAFÉ, Colombia. Informe anual de actividades de la Disciplina de Biometría 2000-2001. Chinchiná: Cenicafé, 2001. 11 p.

OBSERVATORIO NACIONAL DE CIENCIA TECNOLOGÍA E INNOVACIÓN (OCTI). Qué es un indicador? Venezuela, s.f. [Consultado 15 de Diciembre de 2004]. Disponible en Internet: <http://www.octi.gov.ve/indicadores/queson.asp>

OSPINA B., D. Muestreo básico y aplicaciones (parte teórica). En: Simposio de Estadística (3º : 1992 : Bogotá). Muestreo: curso A. Bogotá: Universidad Nacional, 1992. 127 p.

PÉREZ L., C. Técnicas de muestreo estadístico : teoría, práctica y aplicaciones informáticas. México, Alfaomega, 2000. 603p.

PUERTA Q., G. I. Influencia de los granos de café cosechados verdes, en la calidad física y organoléptica de la bebida. En: Cenicafé (Colombia). Vol. 51, No. 2 (2000); p.136-150.

SCHEAFFER, R.; MENDENHALL, W. y OTT, L. Elementos de Muestreo. México, Iberoamericana, 1987. 321 p.

STEEL, R. G. D. y TORRIE, J. Bioestadística: principios y procedimientos. México: McGraw-Hill Book Company, 1988. 662p.

VÉLEZ Z., J. C. La cosecha del café: características y descripción de un método mejorado. En: Curso para caficultores: Cómo mejorar la rentabilidad de las fincas cafeteras, Chinchiná (Colombia), Julio 4 - 7, 2000. Chinchiná: Cenicafé, 2000. 4 p.

VÉLEZ Z., J. C; MONTOYA R., E. C. y OLIVEROS T., C. E. Estudio de tiempos y movimientos para el mejoramiento de la cosecha manual del café. En: Boletín Técnico Cenicafé (Colombia). No. 21 (1999); p. 1-91.

_____. Mejore la recolección de café adoptando el método mejorado. En: Avances Técnicos Cenicafé (Colombia). No. 310 (2003); p. 1-8.

VILLEGAS B., M. J. Influencia de la altura de la planta de café *Coffea arabica* en el desempeño operativo del recolector. Cali, 2003, 65 p. Tesis (Ingeniero Agrícola). Universidad del Valle. Facultad de Ingeniería, 2003.

ANEXO A

ALGORITMO PARA LAS PRUEBAS DE ALEATORIZACIÓN EN SAS

El programa elaborado para realizar las pruebas de aleatorización, es una modificación del presentado por David Cassell para dos muestras independientes[19]. Las modificaciones parten del hecho de que los reordenamientos aleatorios se realizan dentro de divisiones de la totalidad de los datos, y del cambio del estadístico de prueba para este estudio. También debe ajustar la base de datos para el análisis por sitio.

La entrada al programa son 4 bases de datos, cada una representando los valores obtenidos por el muestreo aleatorio simple y un recorrido, en una jornada determinada, para uno de los dos sitios. La salida es una tabla donde se relaciona cada reordenamiento aleatorio con su respectivo estadístico de prueba, incluyendo el estadístico de prueba de los datos originales. Se presenta el caso de la comparación de los valores del indicador eficacia obtenidos por muestreo aleatorio simple y el recorrido recta, en La Soledad.

El código es el siguiente:

[19] CASSELL, D. L. A Randomization-test Wrapper for SAS® PROCs. [Online]. http://www2.sas.com/proceedings/sugi27/p251-27.pdf. (Consultado en Marzo de 2006).

```
%rand_gen(indata=andres.efi_rec_s1,outdata=outrand_1,
    depvar=eficacia,numreps=5000,seed=44380016)

%rand_gen(indata=andres.efi_rec_s2,outdata=outrand_2,
    depvar=eficacia,numreps=5000,seed=19061251)

%rand_gen(indata=andres.efi_rec_s3,outdata=outrand_3,
    depvar=eficacia,numreps=5000,seed=63683583)

%rand_gen(indata=andres.efi_rec_s4,outdata=outrand_4,
    depvar=eficacia,numreps=5000,seed=90919522)


data er1; set outrand_1 outrand_2 outrand_3 outrand_4;
run;

proc sort data=er1; by replicate;
run;

data prescrit;set andres.er1;
run;

proc sort data=prescrit;by replicate metodo jornada;
run;
proc means n noprint data=prescrit;by replicate metodo jornada;var eficacia;
output out=x2 n=n2;
run;

proc sort data=prescrit;by replicate jornada;
run;
proc means mean noprint data=prescrit;by replicate jornada;var eficacia;
output out=x3 mean=m3;
run;

proc sort data=x2;by replicate jornada;
run;
data x4;merge x2 x3;by replicate jornada;
run;

proc sort data=prescrit;by replicate metodo;
run;
proc means n mean noprint data=prescrit;by replicate metodo;var eficacia;
output out=x1 n=n1 mean=m1;
run;
```

```
proc sort data=x1;by replicate metodo;
run;
proc sort data=x4;by replicate metodo;
run;
data x5;merge x1 x4;by replicate metodo;
Pon=n2*m3/n1;
run;

proc sort data=x5;by replicate metodo;
run;
proc means sum noprint;by replicate metodo;var pon;
output out=x6 sum=pon;
run;

proc sort data=x1;by replicate metodo;
run;
proc sort data=x6;by replicate metodo;
run;
data x7;merge x1 x6;by replicate metodo;
S=n1*(m1-pon)**2;
run;

proc sort data=x7;by replicate;
run;
proc means sum noprint;var S;by replicate;
output out=x6 sum=S;
run;
proc print;
run;
```

El programa consta de dos partes:

1. La macro *%rand_gen*, cuyo objetivo es realizar los 5000 reordenamientos aleatorios. Se ejecuta para cada una de las 4 bases de datos que corresponde a los resultados obtenidos para el muestreo aleatorio simple en cada jornada de un sitio determinado. Cada reordenamiento es marcado con un número de 1 a 5000. Los datos originales se rotulan con "0". Luego de ejecutar la macro para cada base de datos se reúnen las 5000 bases de datos resultantes con los

reordenamientos para cada una de las jornadas en una sola base de datos. Esta base de datos es el resultado de realizar 5000 reordenamientos aleatorios dentro de cada jornada.

```
%macro rand_gen(
    indata=_last_,
    outdata=outrand1,
    depvar=y,
    numreps=5000,
    seed=0);

    proc sql noprint;
        select count(*) into :numrecs from
        &INDATA;
        quit;

    data __temp_1;
        retain seed &SEED ; drop seed;
        set &INDATA;
        do replicate = 1 to &NUMREPS;
            call ranuni(seed,rand_dep);
            output;
            end;
        run;

    proc sort data=__temp_1;
        by replicate rand_dep;
        run;

    data &OUTDATA ;
        array deplist{ &NUMRECS } _temporary_ ;
        set &INDATA(in=in_orig)
            __temp_1(drop=rand_dep);
        if in_orig then do;
            replicate=0;
            deplist{_n_} = &DEPVAR ;
            end;
        else &DEPVAR =
            deplist{ 1+ mod(_n_,&NUMRECS) };
        run;

%mend rand_gen;
```

2. Al tener las cuatro bases de datos con los reordenamientos dentro de cada una de las cuatro jornadas, estas se unen en una sola base de datos, en este caso er1, y se organizan de acuerdo al valor de la variable replicate, es decir por el número de reordenamiento. Para cada reordenamiento se calcula el estadístico de prueba planteado, y se obtiene la salida donde se asocia cada reordenamiento al respectivo valor del estadístico de prueba generado. Con esta salida, se debe calcular el porcentaje de los reordenamientos que tienen un estadístico de prueba mayor o igual al obtenido con los datos originales, el cual tiene un valor en la variable replicate = 0.

Este mismo procedimiento, se realizó para la comparación de los valores de los dos indicadores de cada recorrido con el muestreo aleatorio simple, en cada uno de los sitios.

ANEXO B

MAPAS DE CAMPO

Se presentan los marcos de muestreo elaborados para el desarrollo del plan de muestreo probabilístico. El orden es el siguiente:

1. Finca La Soledad, Lote 3A.

2. Finca La Soledad, Lote 5B, partes 1, 2, 3, 4 y 5.

3. Finca La Soledad, Lote 8, partes 1, 2, 3 y 4.

4. Estación Central Naranjal, Lote Castillo 3, parcelas 20, 21, 22, 23 y 24.

Se especifican los surcos en cada terreno. Cada asterisco representa un árbol sembrado en ese sitio. Con la ayuda de estos mapas y reemplazando los asteriscos por números de 1 a *N*, dependiendo del punto de inicio de la jornada de recolección, se realizó la ubicación de los árboles correspondientes a los números aleatorios seleccionados, en los mapas.