



Facultad de Ciencias
Departamento de Estadística

The Survey System: software para la ejecución y gestión de encuestas

Trabajo de titulación presentado por:

Rodrigo Iván Barrera Guajardo

Para optar al título de:

Ingeniero en Estadística

Profesor guía:

Carlos Felipe Henríquez Roldán, PhD

Valparaíso, 2012

Índice general

Índice general	1
Índice de figuras	4
1. Introducción	5
1.1. Encuestas en <i>TSS</i>	6
1.2. Objetivos	9
1.2.1. Objetivos específicos	9
1.3. Hipótesis	9
1.4. Capítulos	9
2. Módulo CATI	11
2.1. Estructura de computadoras	12
2.2. Estructura de archivos del software	12
2.3. Question file	13
2.3.1. Tipos de pregunta de un Question File	13
2.3.2. Tipos especiales de pregunta	14
2.3.3. Crear un Question File en modo batch	15
2.3.4. Ejemplo de Question File creado en modo batch	16

2.3.5. Instrucciones lógicas	17
2.4. Sample File	20
2.4.1. Sample File Specifications	21
2.4.2. Autodiscado	22
2.5. Data File	23
2.5.1. Data File Summary	23
2.6. Captura de voz	24
2.6.1. Control de calidad	25
3. Módulo PDA	27
3.1. Instalar <i>TSS</i> en una PDA	27
3.2. Cuestionarios para PDA	28
3.2.1. Cuestionario off-line para PDA	28
3.2.2. Cuestionario web para PDA	29
3.3. Limitantes de un cuestionario PDA	31
4. Estimación de tamaños de muestra	32
4.1. Simulación	35
4.1.1. Resultados	35
5. Métodos de recolección de datos	39
6. Resultados	42
6.1. Estudio a escala real	43
6.1.1. Ficha técnica	43
6.1.2. Resultados del estudio	44
6.2. Observaciones	49

7. Conclusiones	50
.1. Apéndice	52
.1.1. Do-file simulación	52
.1.2. Abreviaciones	53
Bibliografía	55

Índice de figuras

1.1. Diagrama de flujo de la ejecución de una encuesta en el software <i>TSS</i>	8
2.1. Diagrama de estructura de computadoras	13
2.2. Creando un Question File	14
2.3. Ventana para ingresar instrucciones lógicas	17
2.4. Agregando números al Sample File	21
2.5. Especificaciones del Sample File	21
2.6. Exportando un Data File	23
3.1. Creando un cuestionario PDA off line	29
3.2. Creando un cuestionario web para PDA	30
3.3. Cuestionario web para PDA	30
6.1.	46

Capítulo 1

Introducción

Este trabajo de titulación surgió de la necesidad de contar con una herramienta que permitiera realizar encuestas telefónicas asistidas por computadora prescindiendo de líneas telefónicas convencionales. Esta necesidad promovió la búsqueda de un software que pudiera elaborar y ejecutar instrumentos a través del sistema *CATI* (*Computer Assisted Telephone Interviewing*) y además prescindiera de teléfonos. La búsqueda, llevada a cabo a través de internet, convergió al software *The Survey System* (*TSS*). Éste posee la capacidad de interactuar con Skype, de manera tal de, llamar automáticamente a una serie de teléfonos pre cargados. Por tanto, para implementar el software sólo es necesario un punto de acceso a internet. Implementar el software quiere decir, conectar una serie de computadoras de manera jerárquica, de forma inalámbrica o cableada, para que un conjunto de éstas tengan acceso a un instrumento y canalicen los datos digitados a una computadora central.

Este trabajo apunta, en primer lugar, a profesionales de la estadística y tiene por objeto actualizar los conocimientos sobre metodologías de recolec-

ción de datos. En segunda instancia pretende concientizar a otros profesionales, ligados al análisis de datos, sobre las fases previas a esta etapa. En este sentido se abordará la creación de instrumentos en *TSS* y la estimación de tamaños de muestra a través de simulación. Además se hará un resumen de métodos de recolección de datos y se presentarán los resultados de un estudio a escala real.

1.1. Encuestas en *TSS*

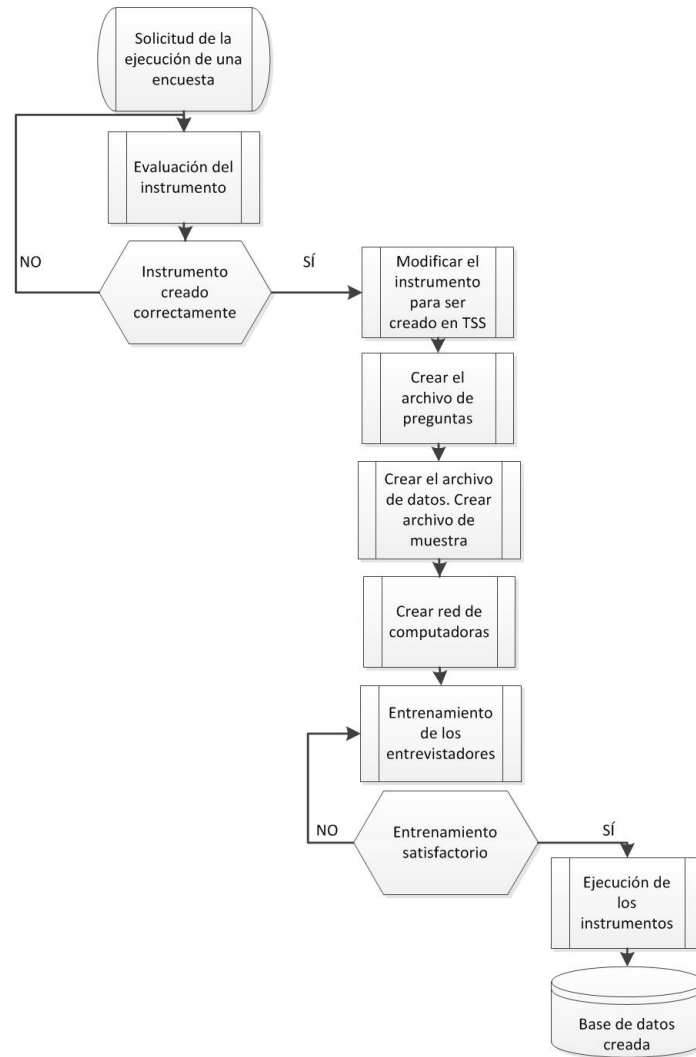
La ejecución de una encuesta en el software *TSS* se puede resumir en los siguientes pasos:

- Perfeccionamiento del instrumento con la contraparte.
- Creación del instrumento en *TSS* en un archivo denominado Question File.
- Entrenamiento de los entrevistadores.
- Exportación de la base de datos.

El tiempo que demore el perfeccionamiento del instrumento está en directa relación con la complejidad de éste. Lo mismo ocurre con la creación del instrumento en *TSS*. El entrenamiento de los entrevistadores es un proceso esencial y se distinguen las etapas de familiarización con el instrumento, ética de trabajo y dinámica con el entrevistado. La dinámica entre entrevistador y entrevistado se puede dar de dos formas, telefónica (a través de Skype) y presencial, a través de PDA (Personal Digital Assistant); ya que, *TSS* permite utilizar estos dispositivos como medio de captura de datos. Para que

TSS pueda ser instalado en una PDA, esta debe tener el sistema operativo Windows CE o Pocket Windows 2002. La exportación de la base de datos es la tarea que permite trasladar los datos obtenidos en *TSS* a un paquete estadístico.

Figura 1.1: Diagrama de flujo de la ejecución de una encuesta en el software TSS



1.2. Objetivos

El objetivo general de este trabajo de tesis es documentar la implementación del software The Survey System considerando aspectos técnicos de la ejecución de encuestas.

1.2.1. Objetivos específicos

- Implementar el módulo *CATI* y PDA del software *TSS*
- Manejar la estructura de archivos de *TSS*.
- Estudiar aspectos técnicos relacionados con la aplicación de encuestas.

1.3. Hipótesis

La hipótesis de este trabajo de tesis es: *The Survey System* gestiona y asiste eficazmente en estudios presenciales y en encuestas telefónicas.

1.4. Capítulos

Este trabajo esta dividido en 7 capítulos, estos son:

- **1 Introducción.**
- **2 Módulo *CATI*.** Se especifica como implementar el módulo *CATI* del software *TSS*. Se hace referencia a los Question File, Data File y Sample File, archivos necesarios para implementar una encuesta telefónica en el software *TSS*.

- **3 Módulo PDA.** Se señala como instalar el software en una PDA y como se crean cuestionarios compatibles con estos dispositivos móviles.
- **4 Estimación de tamaños de muestra.** Se estiman tamaños de muestra, para la distribución multinomial, desde un punto de vista empírico.
- **5 Métodos de recolección de datos.** Se describen los métodos de recolección de datos y sus respectivos soportes.
- **6 Resultados.** Se muestran los resultados de un estudio a escala real.
- **7 Conclusiones.**

Capítulo 2

Módulo CATI

The Survey System es un software que permite recolectar datos de forma telefónica, presencial, a través de correo electrónico y por medio de encuestas web. La recolección de datos de forma telefónica se realiza con el soporte de la tecnología *CATI*. Esta tecnología es ampliamente usada en estudios de opinión, estudios de mercado y estudios políticos. En adición Sturges y Hanrahan (2004) mencionan la utilidad de la tecnología *CATI* en estudios cualitativos.

El sistema *CATI* asiste en forma interactiva a entrevistadores y supervisores en la recolección de datos. Los entrevistadores se sitúan detrás de una terminal (computadora) la cual tiene “cargado” el instrumento que se debe ejecutar. Durante la aplicación del instrumento, *CATI* controla los saltos de pregunta y su respectiva derivación, junto con validar los datos digitados a través de reglas previamente definidas.

El módulo *CATI*, del software *TSS*, cuenta con las siguientes características:

- Posibilidad de implementar las estaciones de entrevista de forma inalámbrica.
- Comunicación telefónica IP o a través de telefonía convencional.
- Creación de filtros, lo que permite que la encuesta tenga una secuencia lógica.
- Fácil exportación de datos a distintos formatos.
- Compatibilidad multimedia para gráficos, videos y audio.

2.1. Estructura de computadoras

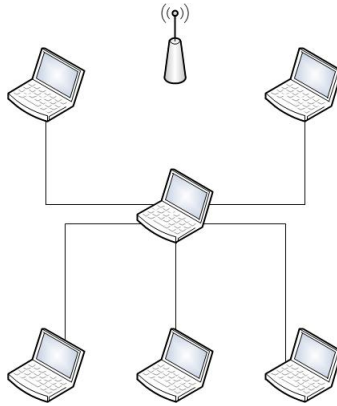
Para implementar y ejecutar un instrumento en el software *TSS* se debe montar una red de computadoras, en la cual existe una computadora central y otras dependientes de esta. Todos los archivos involucrados deben encontrarse en la computadora central, las computadoras dependientes accederán a estos archivos a través de la red; ya sea, en forma de sólo lectura o para sobrescribir en ellos, según sea el caso.

2.2. Estructura de archivos del software

TSS organiza la información del módulo *CATI* en tres archivos principales, estos son:

- **Question file** archivo donde se escribe y configura el instrumento.
- **Sample file** archivo donde se cargan los números de teléfonos.

Figura 2.1: Diagrama de estructura de computadoras



- **Data file** archivo donde se guardan los datos recabados en las entrevistas.

2.3. Question file

Para crear un Question File se debe ir al menú Instructions → Question → New.

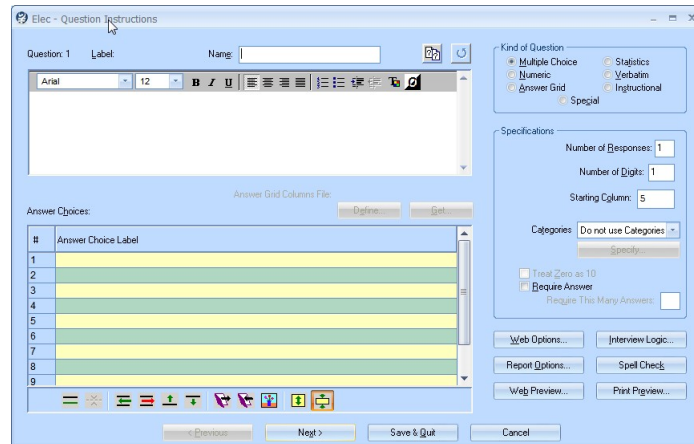
2.3.1. Tipos de pregunta de un Question File

La unidad básica de un Question File son las preguntas. El tipo de pregunta depende del tipo de respuesta que se desea obtener para un tópico en particular.

Los tipos de preguntas son:

- **Multiple Choice:** permite hacer preguntas con una lista predeterminada de respuestas.

Figura 2.2: Creando un Question File



- **Numeric:** permite hacer una pregunta o una serie de preguntas que piden respuestas numéricas (por ejemplo calificaciones o cantidades) en un formato de relleno en blanco.
- **Answer grid:** permite realizar una serie de preguntas relacionadas donde todos tienen la misma lista predeterminada de respuestas y además muestra las posibles respuestas a esas preguntas en un formato de cuadrícula.
- **Verbatim:** preguntas abiertas o comentarios.

2.3.2. Tipos especiales de pregunta

- **Instructional** Es una forma de presentar información o instrucciones al entrevistador.
- **Special** Este tipo de preguntas no aparecen durante el desarrollo de

la encuesta. Sin embargo, pueden registrar variada información en el Data File, por ejemplo:

- ID del entrevistador
- Número de teléfono
- Duración de la entrevista
- Resultado de la llamada

2.3.3. Crear un Question File en modo batch

The Survey System permite crear archivos en procesadores de texto para importarlos y luego convertirlos en Question Files. Una limitante de esta forma de crear Question Files es que no se pueden establecer instrucciones lógicas (ver 2.3.5), sin embargo, estas pueden ser incorporadas en forma interactiva, es decir, mediante una interfaz de usuario.¹

Para realizar lo anterior se deben tener presentes las siguientes consideraciones.

- La primera línea de texto debe ser el título del Question File.
- Cada pregunta debe ser precedida por una Q luego un punto y posteriormente el número de pregunta.
- Para indicarle al software que tipo de pregunta es, se utilizan los Question Code. Los códigos comienzan con [K:, seguidos por Question Kind Code], los que pueden ser:

¹Interfaz de usuario es el medio a través del cual el usuario se comunica con una computadora.

M = Elección múltiple
N = Numérica
A = Grilla de respuesta
V = Literal
I = Instrucción

- Opcionalmente se pueden indicar los siguientes códigos:

ND:# = Número de dígitos
NR:# = Número de respuestas
SC:# = Columna de inicio
RA:# = Respuesta requerida
QN:# = Nombre de la pregunta

2.3.4. Ejemplo de Question File creado en modo batch

Q.1 ¿En qué año nació?

[K:N ND:4]

Año

Q.2 Género

[K:M]

1.Masculino[]

2.Femenino[]

Q.3 ¿Cuán importante son los siguientes ítems para sus estudios?

[K:A]

3 Importante

2 Poco importante

1 Irrelevante

Computadoras última generación

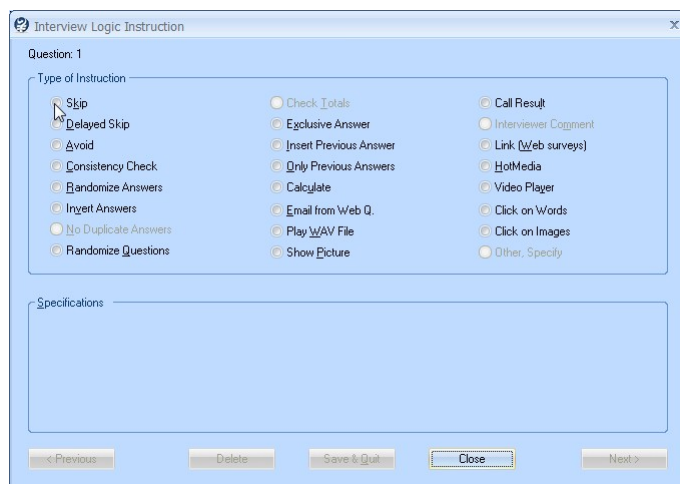
Literatura actualizada

Aula virtual

2.3.5. Instrucciones lógicas

Las instrucciones lógicas son las que dan sentido a los instrumentos. Permiten que a un particular entrevistado, sólo se le realicen las preguntas que corresponden y comprueban que las respuestas sean consistentes. Para poder establecer instrucciones lógicas se debe ir al icono Interview Logic en la ventana del Question File.

Figura 2.3: Ventana para ingresar instrucciones lógicas



Tipos de instrucciones lógicas

- La instrucción Skip le indica a *TSS*, por ejemplo, que salte a la pregunta 5 si la respuesta a la pregunta actual es 1 o 2.
- La instrucción Delayed Skip esperará después que la pregunta actual sea contestada para comprobar la respuesta a una pregunta anterior para ver si un salto debe ejecutarse.
- La instrucción Check Totals suma todas las sub preguntas de una pregunta numérica y compara el total con cierto valor.
- La instrucción Randomize Answers permite aleatorizar la forma en que se leen las posibles respuesta a una pregunta.

Ejemplo: Estableciendo instrucciones lógicas

Instrumento ingresos

Q.1 ¿Cuál es su nombre?

[K:V ND:15]

Q.2 ¿Es usted mayor de edad?

[K:M]

1.Sí

2.No

Q.3 ¿Se encuentra trabajando actualmente?

[K:M]

1.Sí

2.No

Q.4 ¿Gana usted por sobre los 300.000 pesos?

[K:M]

1.Sí

2.No

Q.5 ¿Qué porcentaje de sus ingresos mensuales expende, en promedio, en?

[K:N ND:2]

1.Libros

2.Música

3.Software

Q.6 ¿Cuán importante son los siguientes aspectos en el gasto de sus ingresos?

[K:A]

H Importancia

3 Muy importante

2 Importante

1 No importante

Cultura

Recreación

Alimentación

Q.7

[K:I]

Se agradece su participación. Muchas gracias.

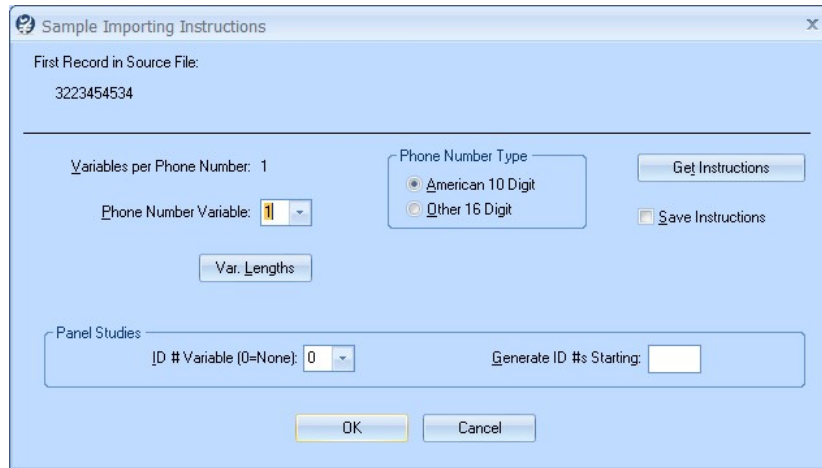
Los saltos que aplican a esta encuesta son:

1. Si Q.1 = No entonces ir a Q.7 (Skip)
2. Si Q.3 = No entonces ir a Q.7 (Skip)
3. Aleatorizar los aspectos de Q.5 (Randomize Answers)
4. Los aspectos de Q.5, sumados deben ser menores a 100 (Check totals)

2.4. Sample File

Sample file es el archivo donde se cargan los números telefónicos. Para incorporar los números en *TSS*, se debe seguir el siguiente procedimiento: En Excel guardar los números telefónicos considerando el código del país y de la región. Luego en *TSS* ir a *Supervise* → *SampleManager* → *Import*

Figura 2.4: Agregando números al Sample File

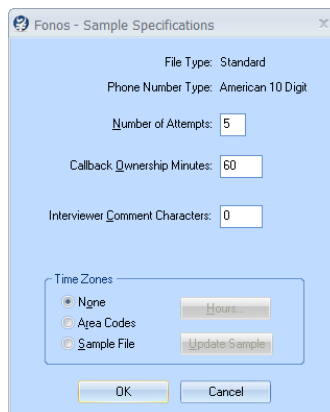


Al hacer clic en OK se despliega la ventana Sample File Specifications

2.4.1. Sample File Specifications

En esta ventana se especifican atributos del Sample File.

Figura 2.5: Especificaciones del Sample File



- **Number of attempts** señala el número de intentos que se realizarán para contactar cada número telefónico.
- **Callback Ownership Minutes** establece número de intentos que se realizarán para contactar a un número telefónico, luego de pactada una cita²

2.4.2. Autodiscado

El módulo *CATI* del software *TSS* permite llamar automáticamente a los números telefónicos señalados en el Sample File.

Métodos de autodiscado

El método de autodiscado se establece en el menú Supervise → Sample Manager → Autodialing Instructions.

- **None** sin autodiscado. El entrevistador deberá marcar el número manualmente.
- **Direct to Modem** The Survey System enviará las instrucciones de autodiscado directamente a los modems.
- **TAPI** (Telephony Application Programming Interface) sistema que permite a windows gestionar la conexión con los modems.

²Las citas son de total relevancia al realizar estudios probabilísticos, ya que es común que la persona seleccionada dentro del hogar para ser entrevistada no se encuentre en el momento en que el entrevistador ejecuta la llamada, por tanto *TSS* registra un intervalo horario donde es posible encontrar al sujeto seleccionado.

- **Skype** este método permite que Skype acceda la base de datos que contiene los números telefónicos para realizar el autodiscado.

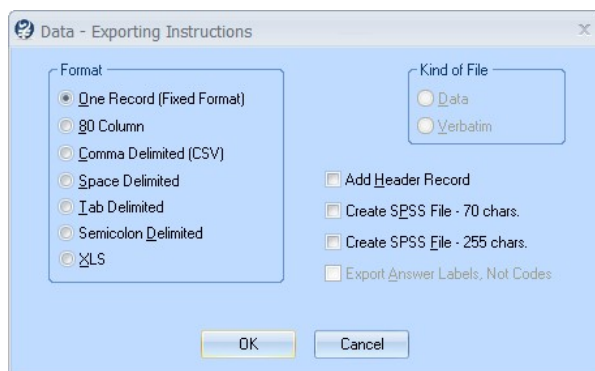
2.5. Data File

TSS almacena las respuestas en un archivos de datos. El mismo archivo de datos se utiliza para entrevistas telefónicas, web o a través de PDA. Este archivo se puede exportar a múltiples formatos.

El sistema almacena los datos de una encuesta en una serie de cuestionarios. Cada cuestionario contiene las respuestas de un sólo individuo.

Para exportar el Data File a formatos legibles por paquetes estadísticos se debe ir al menú Data → Export

Figura 2.6: Exportando un Data File



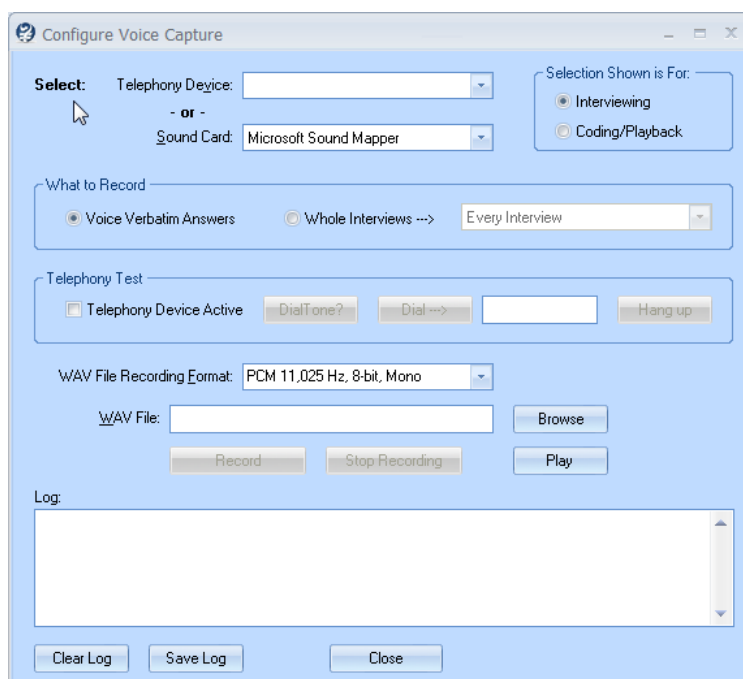
2.5.1. Data File Summary

Permite tener una idea global del contenido del Data File. Este resumen entrega frecuencias y porcentajes. Para acceder al resumen se debe ir al

menú Data → Data File Summary.

2.6. Captura de voz

TSS entrega la posibilidad de grabar las entrevistas . Esta grabación es sin edición, es decir, comienza desde el momento en que el sujeto contesta el teléfono. La configuración de la captura de voz se realiza en el menú Utilities Options.



Las opciones que dispone la captura de voz son: fracción de entrevistas que se grabarán y el formato de audio en que se registrarán.

Oír la voz de un entrevistado posee ventajas respecto a las transcripciones escritas. El tono, la inflexión de la voz y la fuerza de sus opiniones permite

comprender de mejor forma lo que realmente quiso decir. En adición, *TSS* permite codificar las grabaciones de audio en grupos, para luego obtener tablas resumen de dichas codificaciones.

Al ir al menú Data → Code Answers → Inspection se despliega la ventana que permite realizar las clasificaciones.

2.6.1. Control de calidad

Todas las tareas que se ven involucradas en la ejecución de una encuesta telefónica, desde la creación del instrumento hasta la realización de llamadas debe ser supervisada y controlada. En este sentido, Mudryk, Burgess y Xiao (1996) indican los objetivos que debe perseguir la metodología de control de calidad para el sistema *CATI*.

1. Proveer retro alimentación a los operadores de *CATI* en favor del desempeño, esfuerzo personal y mejora.
2. Desarrollar una herramienta para ayudarlos a administrar las operaciones de *CATI* más eficazmente.
3. Proveer una medida estadística objetiva del desempeño y comportamiento del entrevistador.
4. Indicar cuando y donde se requiere entrenamiento adicional.
5. Asegurar el mantenimiento de un alto nivel de ética del trabajo en las operaciones de *CATI*.
6. Proporcionar una garantía de calidad sobre las operaciones de *CATI*.

7. Identificar potenciales areas problemáticas en el diseño del instrumento.

Junto con indicar estos objetivos Mudryk, Burgess y Xiao (1996) establecen medidas y criterios que van en directa mejora de la calidad de los datos que se recolectan con el sistema *CATI*.

Capítulo 3

Módulo PDA

Las PDA se creó con el propósito de ser una agenda electrónica, sin embargo, en el corto plazo, debido al vertiginoso avance tecnológico adoptaron nuevas características y extendieron sus funciones. Los desarrolladores de software vieron el potencial de estos dispositivos y por esto crearon software especializados. Uno de estos software es *TSS*.

3.1. Instalar *TSS* en una PDA

Para que *TSS* puede ser ejecutado en una PDA, primero se debe instalar. Los pasos a seguir para esto son: Hacer clic en el botón **Instalar PDA Software** en el menú del CD de instalación o hacer clic en Setup.msi en la carpeta PDA del CD o bien, copiar el archivo TSSPDA95.armv4.cab dentro de una carpeta de la PDA. Entonces ir a esa carpeta usando el explorador. Pulsar en el archivo.

3.2. Cuestionarios para PDA

TSS permite recolectar datos con PDA de dos formas: off line y basadas en web. El método off line consiste en cargar el instrumento en cada PDA, para luego, una vez completados el número de instrumentos necesarios, traspasar los datos a un computador. El método basado en web, accede al instrumento a través del navegador web de la PDA, por esta razón, no es necesario copiar los archivos en cada dispositivo.

3.2.1. Cuestionario off-line para PDA

Para crear un cuestionario PDA primero se debe que crear un Question File del modo tradicional (ver 2.3), luego ir a PDA → create PDA Questionnaire en el menú principal. Entonces, seleccionar el Question File que se desee utilizar.

El cuestionario PDA cuenta con las siguientes opciones:

- **Automatically Start this Survey** Seleccionar esta opción si se desea que sólo un instrumento este activo y que se inicie de forma automática al abrir la aplicación en la PDA.
- **Automatically Advance to Next Question** Hace que el encuestador no deba pulsar el botón next para ir a la siguiente pregunta
- **Show Question Numbers** La PDA mostrará el número de pregunta al inicio de cada pregunta.

Figura 3.1: Creando un cuestionario PDA off line

The screenshot shows a software window titled "Create PDA Questionnaire". It is divided into several sections for configuring a questionnaire. The "Files to Use" section has a "Question:" input field and a "Browse" button. The "Options" section includes three checkboxes: "Automatically Start this Survey" (unchecked), "Automatically Advance to Next Question" (checked), and "Show Question Numbers" (checked). The "Colors" section allows selecting colors for the "Question Background" (green), "Odd Grid Rows" (cyan), and "Even Grid Rows" (yellow). The "Font" section includes a "Name (blank=default):" field, a "Question Size:" dropdown set to "10", and an "Answer Choice Size:" dropdown set to "10". The "Online Data Collection" section has a "Send Data to a Web Survey" checkbox, a "SSWEB URL:" dropdown menu showing "http://www.surveydemo.com/cgi-bin/", and two text boxes for "Study ID:" and "Participant:". At the bottom of the window are "Create" and "Cancel" buttons.

3.2.2. Cuestionario web para PDA

Los cuestionarios web para PDA son una variación del cuestionario web estándar que es parte del módulo Internet del software *TSS*. Para crear este tipo de cuestionarios, se debe contar con el módulo PDA y el módulo de internet. Los cuestionarios web para PDA están optimizados para el tamaño de la pantalla de estos dispositivos y para las conexiones a internet lentas. Sin embargo, si se desea obviar las limitantes de los cuestionarios para PDA, es posible crear un cuestionario web estándar y acceder a este a través de la PDA.

Para crear un cuestionario web optimizado para PDA, primero se debe

crear un Question File del modo tradicional (ver 2.3) y luego ir al menú PDA
→ Create PDA Web Questionnaire

Figura 3.2: Creando un cuestionario web para PDA

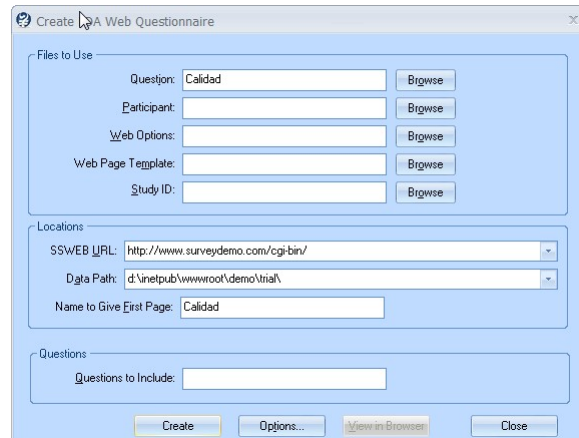
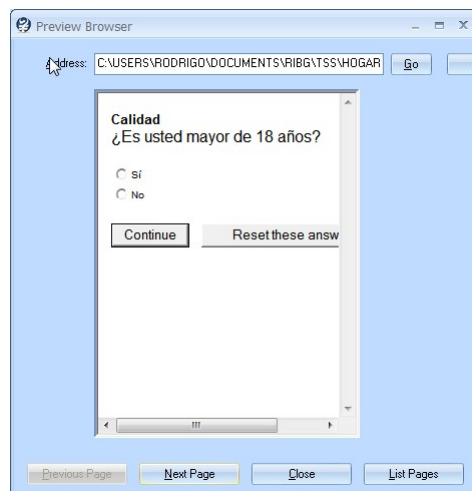


Figura 3.3: Cuestionario web para PDA



3.3. Limitantes de un cuestionario PDA

Las PDA, por ser dispositivos móviles, poseen algunas limitantes en la creación de cuestionarios.

- Las preguntas de opción múltiple no pueden tener más de 99 opciones de respuesta.
- Las preguntas numéricas y las Answer Grid no pueden tener más de 15 sub preguntas.
- Las Answer Grid no pueden tener más de cinco columnas.

Capítulo 4

Estimación de tamaños de muestra

Sin importar la tecnología que se utilice para recolectar datos, se deben estimar tamaños de muestra. En estudios de tipo telefónico, el tamaño de muestra es necesario para determinar la cantidad de números de teléfonos a seleccionar. En estudios cara a cara hay que determinar cuantas personas, hogares, núcleos u otra unidad de estudio deben ser consideradas. El tamaño de muestra afectará directamente la precisión de los estimadores calculados.

En pocas ocasiones existen fórmulas explícitas para calcular tamaños de muestra. En general, en estudios que contemplan más de una variable, el tamaño de muestra debe ser calculado en función de sólo una de éstas; y más precisamente, de la distribución asociada a esta. Un caso excepcional es la distribución binomial. La estimación de tamaños de muestra para esta distribución tiene forma explícita, sin embargo al generalizar esta distribución se obtiene la distribución multinomial, la cual no posee una fórmula que

permita determinar tamaños de muestra.

Angers (1974) estudió el problema de determinación de tamaños de muestra para la distribución multinomial desde una perspectiva gráfica. Thompson (1987) lo abordó utilizando la aproximación normal a la distribución binomial. En este trabajo de tesis el problema se abordará desde la perspectiva de la simulación.

Para calcular tamaños de muestra para una distribución multinomial se debe calcular la probabilidad de que todas las proporciones estimadas estén, simultáneamente, dentro de una distancia especificada, es decir:

$$P\left(\bigcap_{i=1}^k |p_i - \pi_i| \leq d_i\right) = 1 - \alpha \quad (4.1)$$

Thompson (1987) obtiene los siguientes resultados, donde α es el nivel de significación, d es el error muestral y k es la dimensión del vector de parámetros.

Cuadro 4.1: Tamaños de muestra para la distribución multinomial con $d = 0.05$

α	n	m
.50	177	4
.40	203	4
.30	241	3
.20	299	3
.10	403	3
.05	510	3
.025	624	2
.02	664	2
.01	788	2
.005	915	2
.001	1212	2
.0005	1342	2
.0001	1645	2

Fuente: Thompson (1987)

Según Thompson (1987) para obtener un nivel de confianza del 95 % y un error de estimación del 5 % es necesario considerar 510 unidades de estudio. Además para obtener un nivel de confianza del 99 % y un error de estimación del 5 % es necesario considerar 788 unidades de estudio.

4.1. Simulación

Se generó una población de tamaño 100.000 proveniente de una distribución multinomial con vector de parámetros $(\frac{1}{3} \frac{1}{3} \frac{1}{3})$. Sin embargo, como los datos son generados a través de números pseudo aleatorios, el vector de parámetros resultante es (0,33 0,3311 0,3289).

Se extrajeron 100 muestras aleatorias de tamaño $n = 500, 600, \dots, 1.200$, considerando un error de estimación = 0.05, 0.03 y 0.01. Se calcularon niveles de confianza empíricos en función del tamaño de muestra y error de estimación.

4.1.1. Resultados

Cuadro 4.2: Nivel de confianza empírico en función del tamaño de muestra y el error de estimación

Error de estimación	Tamaño muestral							
	500	600	700	800	900	1000	1100	1200
0,05	98,0	98,3	99,6	99,4	100,0	100,0	100,0	100,0
0,03	76,4	78,6	85,3	87,1	87,9	93,1	94,1	94,6
0,01	16,5	20,5	21,0	23,8	25,7	27,8	29,8	33,9

La simulación realizada, determinó que para obtener un nivel de confianza del 95 % y un 3 % de error de estimación, en una distribución multinomial con vector de parámetros (0,33 0,3311 0,3289), es necesario un tamaño muestral

superior a 1200 individuos. Se recalca el hecho que los resultados obtenidos a través de simulación coinciden con los resultados obtenidos por Thompson (1987).

Para extender los resultados se estudiarán tamaños de muestra para poblaciones multinomiales con distintos vectores de parámetros.

Cuadro 4.3: Nivel de confianza empírico en función del tamaño de muestra y el error de estimación. Vector de parámetros (0,3992 0,3020 0,2988)

Error de estimación	Tamaño muestral							
	500	600	700	800	900	1000	1100	1200
0,05	96,5	98,2	99,4	99,0	100,0	100,0	100,0	100,0
0,03	74,7	79,0	83,7	86,4	88,5	93,4	93,0	94,0
0,01	13,6	17,8	20,5	22,8	22,8	27,9	27,9	29,9

Cuadro 4.4: Nivel de confianza empírico en función del tamaño de muestra y el error de estimación. Vector de parámetros (0,6994 0,1010 0,1996)

Error de estimación	Tamaño muestral							
	500	600	700	800	900	1000	1100	1200
0,05	98,5	98,9	100,0	100,0	100,0	100,0	100,0	100,0
0,03	83,4	85,8	92,3	93,6	95,7	96,0	97,6	97,6
0,01	21,4	26,8	28,5	36,1	37,4	40,9	43,5	45,6

Se observa que a medida que los elementos del vector de parámetros están más distantes, el tamaño de muestra para obtener un nivel de confianza, por ejemplo del 95%; con un error de estimación del 3% disminuye. Esta situación es análoga a la que se produce al estimar tamaños de muestra para una distribución binomial. Si no se tiene información sobre la población, se asume $p = 0,5$, ya que en este valor de p la varianza alcanza su máximo. Y, a medida que p se acerca a 0 o a 1 la varianza disminuye, y del mismo modo el tamaño de muestra para obtener cierto nivel de confianza y error de estimación. En adición, si se posee información sobre la población esta debe ser incorporada en el cálculo del tamaño de muestra y, en caso contrario, se debe asumir que la distancia entre los elementos del vector de parámetros es mínima.

Capítulo 5

Métodos de recolección de datos

“El término recopilación de datos es un poco engañoso, ya que implica que los datos ya existen y sólo necesitan ser recogidos. Los datos muestrales son generalmente producidos o creados en el momento de la entrevista o al completar un cuestionario. En otras palabras, son un subproducto del proceso de recolección de datos; un concepto que subraya la importancia de este paso.”

Groves et al (2004)

Los principales métodos de recolección de datos son de forma telefónica, cara a cara y a través de internet (encuestas web). Este último método, generalmente no es recomendable, ya que se pierde el control de la población muestreada. Sin embargo, en encuestas a poblaciones objetivo reducidas en tamaño, la utilidad de este métodos es alta, ya que se pueden aplicar filtros

de manera que sólo el grupo objetivo pueda participar. Los costos monetarios en los que se incurre para implementar este método de recolección son marginales.

En términos comparativos, las encuestas telefónicas se ejecutan en menor tiempo que las encuestas cara a cara. Estas últimas, son las más costosas, sobre todo cuando se abarcan zonas geográficas extensas. En términos de supervisión, son las encuestas telefónicas las que permiten un buen desarrollo de esta tarea, ya que frecuentemente las operadoras telefónicas se encuentran reunidas bajo el mismo espacio físico. Supervisar encuestas cara a cara, además de ser más complejo, es más costoso. Por otro lado, las encuestas cara a cara poseen una mayor interacción con el entrevistado.

La tecnología ha modificado el modo en que estos métodos se ejecutan, creando soportes que van en directa mejora de la calidad e integridad de los datos recolectados. Estas mejoras se ven reflejadas en la disminución de digitación, disponibilidad de grabaciones de audio y facilidad en la creación de cuestionarios. Por ejemplo, las encuestas telefónicas cuentan con los siguientes soportes:

- **Computer-assisted telephone interviewing (CATI)** se genera comunicación entre el entrevistado y el entrevistador telefónicamente, las respuestas son ingresadas automáticamente en una computadora.
- **Interactive voice response (IVR)** telefónicamente se reproduce una grabación con las preguntas y el entrevistado responde utilizando el teclado o bien responde por teléfono y su respuesta es grabada en un archivo de audio.

Los soportes de las encuestas cara a cara son:

- **Computer-assisted personal interviewing (CAPI)** entrevistador y el entrevistado se reúnen, las respuestas son ingresadas automáticamente en una computadora.
- **Audio computer-assisted self-interviewing** (audio-CASI or ACASI) el entrevistado opera un computador, este muestra en la pantalla las preguntas. Las respuestas son grabadas en un archivo de audio.
- **PDA** el instrumento es diseñado en una computadora y luego cargado en una PDA.
- **OCR** reconocimiento óptico de caracteres
- **OMR** reconocimiento óptico de marcas

Se debe señalar que no es correcto preferir un método por sobre otro taxativamente, sino que deben ser considerados aspectos de costos, tiempos de ejecución y pregunta de investigación. En este sentido Bonnel (2001) señala que no es posible concluir que un método de recolección sea superior a los demás. Cada uno tiene ventajas y desventajas. Además, las características culturales o nacionales, limitan la aplicabilidad de ciertos tipos de métodos de recolección. En adición, Sturges y Hanrahan (2004) comparan los estudios cualitativos ejecutados de forma cara a cara y telefónicamente concluyendo que las entrevistas telefónicas pueden ser utilizadas con éxito en estudios cualitativos.

Capítulo 6

Resultados

Una vez implementado el software se realizaron diversas pruebas, incluyendo un estudio a escala real, donde se puso en práctica el diseño de instrumentos, se extrajeron números telefónicos al azar y se obtuvo una base de datos. Durante la implementación se crearon instrumentos en modo batch y en forma interactiva y se logró dominar los aspectos técnicos sujetos a esta tarea. También fue posible establecer metodologías de trabajo que, una vez exportados los datos a un paquete estadístico, permiten incorporar el plan de muestreo en los análisis.

Sobre la base de lo anterior se comprobó que, en efecto, *TSS* es un software completo en términos de que permite recolectar datos de forma telefónica, cara a cara, utilizando correo electrónico y a través de encuestas web. La creación de cuestionarios complejos es una tarea que se puede realizar de manera sencilla y efectiva, destacando que los cuestionarios se crean del mismo modo, sin importar la plataforma que se utilice para recolectar los datos. El software asiste eficazmente en la ejecución de estudios de distinta índole,

siendo una herramienta valiosa en la recolección de datos, junto con aportar significativamente a la calidad de los mismos, debido a que cuenta con diversas herramientas de control y revisión, como por ejemplo generación de números aleatorios, grabaciones de audio y metadatos.

El ambiente de programación de *TSS* no existe como tal, sino que se pueden crear instrumentos en procesadores de texto para luego importarlos. Este es un aspecto en el cual el software presenta deficiencias, ya que no se pueden programar las encuestas en su totalidad debido a que las instrucciones lógicas sólo se pueden establecer de forma interactiva.

6.1. Estudio a escala real

El estudio a continuación detallado fue ejecutado por el Centro de Estudios Estadísticos de la Universidad de Valparaíso. Es un estudio de intención de voto y, para mantener la ética estadística, se mantendrán en reserva el nombre de los candidatos y la comuna donde tuvo lugar. El software *TSS* fue utilizado para realizar las entrevistas telefónicas y el marco muestral fue el software Green CD. Este software posee la base de datos de todos los teléfonos de Chile, discriminando por hogares y empresas.

6.1.1. Ficha técnica

- **Población objetivo:** habitantes inscritos en los registros electorales de la comuna.
- **Población muestreada:** habitantes que disponen de telefonía de red fija inscritos en los registros electorales de la comuna.

- **Tamaño muestra:** 450 habitantes inscritos en la comuna. Error de estimación 4% y un nivel de confianza del 95%.
- **Tipo de estudio:** encuesta telefónica probabilística. Las llamadas fueron efectuadas los días 14 y 15 de diciembre de 2011.

6.1.2. Resultados del estudio

Sexo:	Edad promedio
Femenino	54.7
Masculino	52.4

La edad promedio de los entrevistados está acorde al padrón electoral chileno envejecido.

1. Si las elecciones para alcalde fueran el próximo fin de semana, ¿por quién votaría usted? (Sin mencionar candidatos)

Intención de voto:	Porcentaje
Otro	1.07
Candidato 1	8.37
Candidato 2	29.36
Candidato 3	0.07
NS/NR	60.90
Votará NULO	0.23
Total	100.00

Cabe destacar que la mayoría de las personas entrevistadas no sabe o no responde de manera espontánea por quién votaría en una eventual elección de alcalde que se realizará el próximo fin de semana.

2. Si en la próxima elección de alcalde en la comuna de _ compitieran los siguientes cuatro candidatos, ¿por cuál de ellos votaría usted?

Intención de voto:	Porcentaje
Candidato 1	15.71
Candidato 2	34.03
Candidato 3	0.17
NS/NR	48.39
Votará BLANCO	1.47
Votará NULO	0.23
Total	100.00

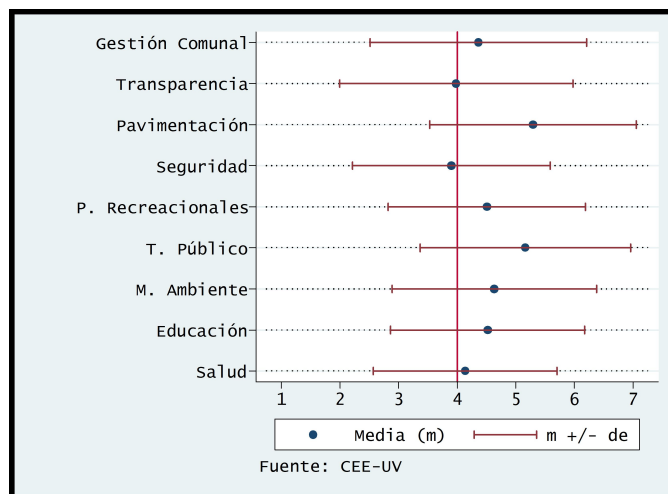
3. En un escenario diferente de candidatos. Si en la próxima elección compitieran las siguientes personas, ¿por cuál de ellos votaría?

Intención de voto:	Porcentaje
Candidato 1	14.27
Candidato 2	1.47
Candidato 3	35.40
Candidato 4	0.07
NS/NR	48.39
Votará BLANCO	0.10
Votará NULO	0.30
Total	100.00

Cabe destacar que al nombrar una cantidad cerrada de candidatos a alcalde el porcentaje de quienes no saben o no responden disminuye drásticamente.

4. Evalúe al actual alcalde -en una escala de 1 a 7- en los siguientes aspectos:

Figura 6.1:



Prácticamente, todos los aspectos evaluados están con notas cercanas a 4,0. El aspecto con mejor evaluación es la pavimentación que supera levemente la nota 5,0.

5. De los aspectos antes evaluados cuáles son los más prioritarios para usted. Se señalan las primeras tres prioridades.

Prioridad 1:	Porcentaje

Salud	45.55
Educación	13.06
M. Ambiente	8.67
T. Público	5.36
P. Recreacionales	21.30
Seguridad	3.15
Pavimentación	1.17
Transparencia	1.74

Total	100.00

Prioridad 2:	Porcentaje

Salud	20.87
Educación	30.18
M. Ambiente	3.76
T. Público	2.29
P. Recreacionales	30.60
Seguridad	4.08
Pavimentación	0.18
Transparencia	8.03

Total	100.00

Prioridad 3: Porcentaje	
Salud	7.40
Educación	19.57
M. Ambiente	12.92
T. Público	4.49
P. Recreacionales	37.17
Seguridad	6.93
Pavimentación	8.33
Transparencia	3.18
Total	100.00

Llama extremadamente la atención la alta prioridad que le otorgan los entrevistados a los parques recreacionales.

6. Sin considerar el orden se indica el porcentaje de entrevistados que menciona el aspecto.

Aspecto	Sí (%)	No (%)
Salud	63.5	36.5
Educación	43.5	56.5
Medio ambiente	17.1	82.9
Transporte público	11.2	88.8
Parques recreacionales	57.1	42.9
Seguridad	10.2	89.8
Pavimentación	5.9	94.1
Transparencia	11.3	88.7

Esto se ve reflejado al no considerar las prioridades.

6.2. Observaciones

Para que una encuesta se considere probabilística, se debe también seleccionar al azar al individuo que contestará dentro de la vivienda. Se notó que este proceso reduce la tasa de respuesta de forma drástica. Esto se debe a que el mecanismo de selección al azar dentro de la vivienda requiere interacción con el entrevistado.

Otro factor influyente en la tasa de respuesta es el tipo de estudio, por ejemplo, los estudios de tipo político afectan la tasa de respuesta en forma negativa.

Capítulo 7

Conclusiones

Se logró implementar exitosamente el software *TSS*, tanto el módulo *CATI* como el módulo PDA, sin embargo, es necesario precisar que los dispositivos móviles PDA están prácticamente obsoletos debido al rápido avance de la tecnología en dispositivos similares (smartphones). Los desarrolladores de software están por detrás de la tecnología actual respecto a la utilización de dispositivos móviles para la recolección de datos. Para que *TSS* pueda ser instalado en una PDA, esta debe tener el sistema operativo Windows CE o Pocket Windows 2002, los cuales están obsoletos. Sin embargo Microsoft desarrolló otro sistema operativo para dispositivos móviles, pero no es posible determinar si el software *TSS* será compatible con dicho sistema operativo ya que, no fue posible, en el transcurso de este trabajo de titulación, acceder a un dispositivo que contara con este sistema.

El software *TSS* asiste en forma eficaz en la recolección de datos de forma telefónica, pero este proceso debe ser estrictamente supervisado a través de: creación de estadísticas por entrevistador, grabación de entrevistas y

metadatos. Es necesario además, la capacitación de los entrevistadores y la retroalimentación de los mismos.

El software *TSS* se implementó en su totalidad en forma inalámbrica. Se experimentaron ocasionalmente deficiencias en la red cuando el número de computadoras era superior a seis. Por esto se recomienda montar una red cableada, ya que esta es más estable. Sin embargo, el software funciona normalmente, de forma inalámbrica, con menos de seis computadoras.

La aleatoriedad, parte fundamental de la estadística, debe estar presente en todo el proceso de recolección de datos de forma telefónica, comenzando con la elección de números telefónicos y, finalizando con la elección de sujetos dentro del hogar. Se comprobó que las tasas de respuesta se reducen significativamente al realizar selección aleatoria dentro del hogar, núcleo u otra unidad de estudio. No obstante, es este procedimiento el que permite incorporar, de forma correcta, los factores de expansión en el análisis de datos.

La construcción de instrumentos en el sistema *CATI* debe ser sumamente cuidadosa, presentando atención a la lógica de estos. Además, se deben entregar mensajes instructivos para los entrevistadores en las preguntas y hacer énfasis en que la lectura de estas sea exacta, evitando su simplificación. Se observó que los entrevistadores tienden a simplificar los instrumentos una vez que se familiarizan con este.

.1. Apéndice

.1.1. Do-file simulación

Código que permite calcular tamaños de muestra para una población multinomial a través de simulación. Este programa fue escrito en el software Stata, versión 11.

```
clear all
cd"C:\Users\Documents\"
mata
rseed(130)

p = (0.4,0.3,0.3)
votos = rdiscrete(100000,1,p)
end

getmata votos
save votos, replace
preserve
contract votos, percent(p)
list p
restore

set more off

capture program drop n
program multinomial
mata IC = J(10, 1, 0)
forvalues j = 1(1)10{
local ic = 0
```

```

forvalues i = 1(1)100{
    clear
    use votos.dta
    quietly sample '1', count
    contract votos, percent(p)
    if p[1] > 40.92 | p[1] < 38.92 | p[2] > 31.20 | p[2] < 29.20 {
        local ic = 'ic'
    }
    else{
        local ic = 'ic' + 1
    }
}
display 'ic'
mata IC['j',] = 'ic'
}
end
}

```

.1.2. Abreviaciones

- **ACASI** Audio computer-assisted self-interviewing
- **CAPI** Computer-assisted personal interviewing
- **CATI** Computer Assited Telephone Interviewing
- **IVR** Interactive voice response
- **OCR** Optical Character Recognition
- **OMR** Optical Mark Recognition

- **PDA** Personal Digital Assistant
- **TSS** The Survey System

Bibliografía

- [1] Angers, C. (1974). A Graphical Method to Evaluate Sample Sizes for the Multinomial Distribution. *Technometrics*, Vol. 16, No. 3, pp. 469-471
- [2] Bonnel, P. (2001). *Transport Survey Quality and Innovation*. Postal. Telephone and face to face surveys: how comparable are they. UK, Oxford: Elsevier Science Ltd.
- [3] Fernández, I. (2005). Encuestas telefónicas asistidas: Implementación del sistema informático Blaise CATI . Trabajo de titulación para optar al título de ingeniero estadístico. Universidad de Valparaíso, Chile.
- [4] Groves, R., Fowler, J., Couper, M., Lepkowski, J., Singer, E. Tourangeau, R. (2004) *Survey methodology*, NY, New York: John Wiley & Sons.
- [5] Groves, R. M. (1990). Telephone and methods of telephones surveys. *Annual Review of Sociology*. Vol. 16, pp. 221-240
- [6] Heeringa, S. G., West B. T. y Berglund P. A. (2010). *Applied Survey Data Analysis*. FL, Boca Raton: CRC Press.

- [7] House, C. C. (1985). Questionnaire Design With Computer Assisted Telephone Interviewing. *Journal of Official Statistics*, Vol.1, No.2, pp. 209–219
- [8] Mudryk, W. M., Burgess, J. y Xiao, P. (1996). Quality control of CATI operations in statistics. *Proceedings of the section on Survey Research Methods*. American Statistical Association.
- [9] Blumberg, S. J., Luke, J. V., Cynamon, M. L. (2006). Telephone Coverage and Health Survey Estimates: Evaluating the Need for Concern About Wireless Substitution. *American Journal of Public Health*. Vol 96, No. 5. 926-931.
- [10] Sturges, J. E. y Hanrahan, K. J. (2004). Comparing Telephone and Face-to-Face Qualitative Interviewing: A Research Note. *Qualitative Research* Vol 4, pp. 107
- [11] Thompson, S. K. (1987). Sample Size for Estimating Multinomial Proportions. *The American Statistician*, Vol 41(No.1), pp. 42-46.
- [12] Tortora, R. D. (1978). A Note on Sample Size Estimation for Multinomial Populations. *The American Statistician*, Vol 32, pp. 100-102.