



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

UNIVERSIDAD DE LA REPÚBLICA

FACULTAD DE CIENCIAS ECONÓMICAS Y DE
ADMINISTRACIÓN

TRABAJO FINAL DE GRADO

metaSurvey

Paquete de R para el procesamiento de encuestas por muestreo con
generación de recetas mediante metaprogramación y estimación de
varianzas.

Estudiante:
Mauro Loprete

Tutora:
Dra. Natalia da Silva

Trabajo final de grado presentado como requisito para la obtención
del título Licenciado en Estadística

Resumen

metaSurvey

por Mauro Loprete

El trabajo presenta metaSurvey, un paquete de R diseñado para mejorar el procesamiento de encuestas por muestreo y la estimación de parámetros poblacionales. Utiliza metaprogramación y técnicas de remuestreo para ofrecer resultados precisos, evaluar la incertidumbre y fomentar la reproducibilidad. A diferencia de otras bibliotecas, metaSurvey combina flexibilidad mediante metaprogramación con las capacidades de procesamiento de encuestas del paquete survey. Los objetivos incluyen proporcionar una herramienta útil, incorporar técnicas de remuestreo para usuarios no expertos, permitir la generación de ‘recetas’ personalizadas, y fomentar la contribución de la comunidad. Se destaca como alternativa a paquetes propietarios, enfocándose en la transparencia y reproducibilidad para mejorar la confiabilidad de las estimaciones poblacionales.

Agradecimientos

Acá le voy a agradecer a alguien?

Índice

Resumen	iii
Agradecimientos	v
1 Introducción	1
1.1 Motivación	2
1.2 Mercado de Trabajo {ech} vs {metaSurvey}	3
2 Metaprogramación	5
2.1 Principios básicos	5
2.1.1 El código cmo datos	5
2.1.2 El código como expresiones	6
2.1.3 El código puede generar código	6
2.1.4 El código puede ser evaluado	6
3 Muestreo	7
4 Metodología de varianzas	9
5 Desarrollo del paquete	11
6 Infraestructura	13
7 Aplicación	15
8 Conclusiones y trabajo a futuro	17
Appendices	19
Soy un apaendice	19

Figuras

Tablas

Acá va la dedicatoria

Chapter 1

Introducción

En este documento se presenta el desarrollo del paquete `metaSurvey` disponible en R. El objetivo principal de `metaSurvey` es permitir que el usuario pueda construir indicadores de manera reproducible y transparente, teniendo el usuario un control total sobre el proceso de transformación de los microdatos a indicadores.

A lo largo del documento se mencionan diferentes conceptos para el desarrollo del paquete, como la meta-programación, conceptos de Inferencia en Poblaciones finitas, metodología sobre el desarrollo de paquetes en R. Además, se presentan ejemplos de cómo utilizar el paquete `metaSurvey` para construir indicadores de mercado de trabajo a partir de los microdatos de la Encuesta Continua de Hogares del Instituto Nacional de Estadística de Uruguay.

La motivación principal del desarrollo del paquete fue la necesidad de contar con un paquete que permita al usuario tener un control total sobre el proceso de transformación de los microdatos a indicadores. En general, los paquetes que existen en R para el análisis de encuestas por muestreo son muy sensibles a la estructura y las variables que componen a la encuesta. En general, un cambio en la estructura de la encuesta implica una nueva versión del paquete utilizado para obtener los indicadores, siendo poco flexible a cambios en la estructura de la encuesta y generalmente el usuario tiene una función de alto nivel que no permite modificar el código para adaptarlo a sus necesidades o saber cada paso que se realiza para obtener el indicador sin tener que leer el código fuente o la documentación adjunta.

El problema de sensibilidad a la estructura de la encuesta puede verse en el paquete `ech` ([Mathieu & Detomasi, 2020](#)) donde en el existen funciones para crear variables de mercado de trabajo, pero estas funciones dependen de una va

Dentro del documento se presentan ejemplos de cómo utilizar el paquete `metaSurvey` para construir indicadores de mercado de trabajo a partir de los microdatos de la Encuesta Continua de Hogares del Instituto Nacional de Estadística de Uruguay.

Además, se presentan conceptos claves de la meta-programación y se muestra cómo se utiliza la meta-programación en el paquete `metaSurvey`. Luego se presentan conceptos básicos de muestreo sobre los estimadores bajo el contexto de Inferencia en Poblaciones finitas, para luego mostrar cómo es crucial tener en cuenta el diseño de la encuesta para la construcción de indicadores y evaluar su incertidumbre, otro de los puntos fuertes del paquete `metaSurvey`. Por último, se muestra una aplicación del uso cotidiano del paquete, donde pueden verse ejemplos de cómo construir nuestras propias recetas y/o también consultar recetas ya construidas por otros usuarios.

1.1 Motivación

La fuente principal de los indicadores sociales y económicos a nivel poblacional proviene, en general, de encuestas por muestreo. Construir un indicador de interés a partir de los microdatos de una encuesta implica un proceso de creación de variables que intentan capturar el fenómeno de interés. En general, el proceso de construcción de indicadores es complejo e involucra la selección de variables, la creación de variables sintéticas, la imputación de valores faltantes, la estandarización de variables, la selección de la unidad de análisis, entre otros.

Este proceso es crucial para la correcta interpretación de los resultados y para la comparabilidad de los indicadores a través del tiempo y entre diferentes poblaciones. En este sentido, la construcción de indicadores requiere un conocimiento profundo de la encuesta y de las variables que la componen, así como de las características de la población de interés.

En general, este proceso se realiza en algún software estadístico, como SPSS, SAS o STATA. Sin embargo, en los últimos años, el uso de R para el análisis de datos ha crecido significativamente. R es un lenguaje de programación y un entorno de desarrollo para la realización de análisis estadísticos. R es un software libre, lo que significa que es gratuito y actualmente posee una comunidad de usuarios muy activa que desarrolla paquetes y funciones para realizar análisis y modelos estadísticos.

En R, existen diferentes paquetes para procesar encuestas por muestreo. El paquete más conocido es `survey`, que permite realizar análisis de encuestas por muestreo de manera sencilla y eficiente. El paquete `survey` permite realizar análisis descriptivos, estimar diferentes indicadores tomando en cuenta el diseño de la encuesta y los estadísticos más usuales. Además, el paquete `survey` permite realizar análisis de encuestas complejas, como encuestas con estratificación, conglomerados y diseño de muestreo complejo.

Dentro de la comunidad de R también existen otros paquetes que se centran en alguna encuesta en particular, como el paquete `ech`, que permite obtener datos de la Encuesta Continua de Hogares del Instituto Nacional de Estadística de Uruguay, o el paquete `eph`, que permite obtener datos de la Encuesta Permanente de Hogares del Instituto Nacional de Estadística y Censos de Argentina, entre otros.

Estos paquetes permiten democratizar el acceso a los microdatos y obtener los indicadores de interés utilizando R con funciones comprensibles para cualquier usuario con experiencia en análisis de datos. Esto es importante ya que permite que los usuarios puedan replicar los resultados y verificar la calidad de los indicadores. Además, fomenta la transparencia y la reproducibilidad de los resultados, ya que los usuarios pueden compartir el código y los datos utilizados para obtener los indicadores.

Sin embargo, una de las limitaciones de estos paquetes es que son muy sensibles a la estructura y las variables que componen a la encuesta. En general, un cambio en la estructura de la encuesta implica un cambio en el código utilizado para obtener los indicadores, siendo poco flexible a cambios en la estructura de la encuesta y generalmente el usuario tiene una función de alto nivel que no permite modificar el código para adaptarlo a sus necesidades o saber cada paso que se realiza para obtener el indicador sin tener que leer el código fuente o la documentación adjunta.

En este sentido, `metaSurvey` es un paquete que busca solucionar estas limitaciones ya que todo el proceso de transformación de los microdatos a indicadores se realiza a

través de una serie de funciones que permiten al usuario tener un control total y transparente sobre el proceso de transformación de los microdatos a indicadores. Además, metaSurvey permite que el usuario pueda realizar el proceso de transformación de los microdatos a indicadores de manera reproducible y transparente, ya que el usuario puede compartir el código y los datos utilizados para obtener los indicadores, mediante lo que denominamos steps y recipes, conformando así una especie de “recetario de cocina” para la construcción de indicadores.

El enfoque que permite la flexibilidad a la hora de construir los indicadores es la meta-programación. La meta-programación es un paradigma de programación que permite que un programa pueda modificar su estructura interna. En R, la meta-programación se realiza a través de las funciones eval y parse que permiten evaluar y parsear código de manera dinámica. En este sentido, metaSurvey utiliza la meta-programación para permitir que el usuario pueda modificar el código que se utiliza para transformar los microdatos a indicadores teniendo funciones similares a las que se utilizan en el paquete recipes de la librería tidymodels.

En lo que sigue de esta sección, se presenta el paquete metaSurvey y se muestran ejemplos para la construcción y estimación de variables de mercado de trabajo de Uruguay con la ECH 2022 utilizando el paquete metaSurvey y el paquete ech.

En capítulos posteriores se presentan conceptos claves de la meta-programación y se muestra cómo se utiliza la meta-programación en el paquete metaSurvey. Luego se presentan conceptos básicos de muestreo sobre los estimadores bajo el contexto de Inferencia en Poblaciones finitas, para luego mostrar cómo es crucial tener en cuenta el diseño de la encuesta para la construcción de indicadores y evaluar su incertidumbre, otro de los puntos fuertes del paquete metaSurvey. Por último, se muestra una aplicación del uso cotidiano del paquete, donde pueden verse ejemplos de cómo construir nuestras propias recetas y/o también consultar recetas ya construidas por otros usuarios

1.2 Mercado de Trabajo {ech} vs {metaSurvey}

Chapter 2

Metaprogramación

En este capítulo se introducen los conceptos de Metaprogramación en R y se presentan las funciones básicas para la manipulación de expresiones y funciones así como también haciendo un breve mención sobre la aplicación en **metaSurvey**.

2.1 Principios básicos

Dentro de la programación existen diferentes paradigmas de programación, entre ellos se encuentra la programación orientada objetos (OOP), la programación funcional, programación lógica entre otras. En R, el paradigma predominante es la programación orientada a objetos, sin embargo, R también permite la programación funcional y la programación lógica, la facilidad o aplicación depende del contexto y la utilización de ciertas librerías. La metaprogramación es una técnica que se puede aplicar en cualquier paradigma de programación, sin embargo, en R la metaprogramación se aplica principalmente en la programación funcional.

La metaprogramación es una técnica que permite que el código pueda ser manipulado como datos. En R, el código se representa como una expresión. Una expresión es una estructura de datos que representa una instrucción o una parte de una instrucción. Por ejemplo, la expresión `x <- 1` representa la instrucción de asignar el valor 1 a la variable `x`. La expresión `x + 1` representa la instrucción de sumar 1 al valor de la variable `x`. La expresión `x <- 1; x + 1` representa la instrucción de asignar el valor 1 a la variable `x` y luego sumar 1 al valor de la variable `x`. Una expresión puede ser tan simple como una variable o tan compleja como una función anónima (o también conocida como función lambda).

En R, el código es un tipo de dato más. Esto permite que el código pueda ser manipulado como cualquier otro tipo de dato. Por ejemplo, se puede asignar código a una variable, pasar código como argumento a una función, o devolver código como resultado de una función. Manipular al código como datos es lo que se conoce como metaprogramación.

Para entender la metaprogramación es necesario entender los siguientes conceptos:

2.1.1 El código como datos

Dentro de este paradigma, no vamos a operar con un conjunto de datos rectangular sino que vamos a operar con los argumentos de una función (**formals**), el cuerpo de una función (**body**), el entorno de una función (**environment**), para generar de forma dinámica y flexible código para manipular, procesar y aplicar inferencia sobre encuestas por muestreo.

2.1.2 El código como expresiones

Al operar con expresiones es importante saber el detrás de escenas al ejecutar código este concepto es conocido como **Abstract syntax tree**:

```
# TODO: Cambiar el ejemplo por uno de metaSurvey haciendo referencia al engine.  
lobstr::ast(f(a, "b"))  
#> f  
#> a  
#> "b"
```

2.1.3 El código puede generar código

2.1.4 El código puede ser evaluado

Chapter 3

Muestreo

Chapter 4

Metodología de varianzas

Chapter 5

Desarrollo del paquete

Chapter 6

Infraestructura

- Infra
- Docker
- Kubernetes
- Tests
- Envío a CRAN

Chapter 7

Aplicación

- Mostrar Series Históricas
- Aplicación ANII
- Compartir recetas

Chapter 8

Conclusiones y trabajo a futuro

- Mostrar Series Históricas
- Aplicación ANII
- Compartir recetas

Soy un apaendice

Mathieu, G., & Detomasi, R. (2020). *Ech: Caja de herramientas para procesar la encuesta continua de hogares*. <https://github.com/calcita/ech>