

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Paquetes para el tratamiento de datos provenientes de encuestas aleatorizadas

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Beatriz Cobo Rodríguez
María del Mar Rueda García



beacr@ugr.es, mrueda@ugr.es
Departamento de Estadística e Investigación Operativa
Universidad de Granada

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

1 Introducción

2 Paquetes R

- Paquete RRTCS
 - Modelos cualitativos
 - Modelos cuantitativos
- Paquete RRreg
- Paquete rr
- Paquete list
- Paquete endorse

3 Bibliografía

Introducción

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Estudios de encuestas por muestreo

↓ *Interés*

Aspectos sensibles o confidenciales

↓ *Problema*

Deseabilidad social: {Tendencia de las personas encuestadas a responder en función de lo que es aceptable socialmente}

↓

Rehúsan
participar encuesta

Respuestas
falsas

Respuestas
condicionadas

↓

Alteran la precisión y confiabilidad de los estimadores

Introducción

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Solución



Técnica de Respuestas Aleatorizadas (Warner, 1965)



Protección
del
anonimato

Reducir
el riesgo de evasión o
no respuesta

- ¿Usted consume drogas?
- ¿Usted no consume drogas?

Introducción

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ventajas:

- Aumenta la probabilidad de contestar la verdad respecto a una pregunta directa.
- Mayor índice de respuesta.

Desventajas:

- Aumento en la complejidad de la pregunta.
- Dificultad en entender el método de aleatorización.
- Requiere de muestras de tamaños grandes.

Métodos de respuesta aleatorizada

↓ propuestos

Muestreo aleatorio simple

Encuestas

↓ se realizan

Diseños muestrales complejos: estratos, conglomerados...

↓ requieren

Un tratamiento diferenciado derivado de la no independencia de las unidades muestrales.

⇓

CRAN contiene algunos paquetes de R (survey, sampling, laeken or TeachingSampling) que incluyen los métodos usados en estimación en muestreo.

↓

No pueden usarse directamente cuando la muestra es obtenida mediante respuesta aleatorizada.

Paquetes R

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Muy recientemente algunos autores han desarrollado paquetes de R para la estimación con respuesta aleatorizada o procedimientos similares.

Estos paquetes son:

- Paquete RRTCS
- Paquete RRreg
- Paquete rr
- Paquete list
- Paquete endorse

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Descripción

El paquete RRTCS: Randomized Response Techniques for Complex Surveys (Cobo, Rueda y Arcos, 2015)

- Realiza estimaciones puntuales y por intervalos para totales, medias y proporciones con datos obtenidos de encuesta de respuesta aleatoria.
- Trabaja con una amplia gama de diseños de muestreo incluyendo muestreo aleatorio simple con y sin reemplazamiento (SRSWR, SRSWOR), muestreo estratificado, muestreo por conglomerados, muestreo con probabilidades desiguales y combinaciones de ellos.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción

- Se compone de veinte funciones principales, cada una de ellas implementando uno de los procedimientos de RR para encuestas complejas.
- Tiene una función adicional llamada *Resampling Variance* que proporciona estimaciones de la varianza de los estimadores de respuesta aleatoria utilizando algunos métodos de remuestreo.
- Incluye 20 conjuntos de datos que contienen observaciones de de diferentes encuestas realizadas en poblaciones reales y simuladas utilizando diferentes técnicas de respuesta aleatoria.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

**Modelos
cualitativos**

Modelos
cuantitativos

Paquete RRreg

Paquete list

Paquete
endorse

Bibliografía

Ejemplo

Nuestro objetivo es estimar la prevalencia de estudiantes que consumen drogas.

Para ello:

- extraemos una muestra mediante un mecanismo de muestreo
- aplicamos las distintas técnicas cualitativas de respuesta aleatoria disponibles en el paquete RRTCS

Paquete RRTCS

Warner

El mecanismo de respuesta aleatoria de Warner funciona del siguiente modo:

- A una persona de la muestra etiquetada i se le ofrece una caja con un número considerable de cartas idénticas
- con una proporción $p(0 < p < 1; p \neq 0.5)$ de ellas marcadas *¿Consumes drogas?* y el resto marcadas *¿No consumes drogas?*.
- Se le pide a la persona que extraiga aleatoriamente una de ellas, observe la marca de la carta y de una respuesta

$$z_i = \begin{cases} 1 & \text{si la carta está marcada "¿Consumes drogas?"} \\ & \text{y de verdad "Consumes drogas", o} \\ & \text{si la carta está marcada "¿No consumes drogas?"} \\ & \text{y de verdad "No consumes drogas"} \\ 0 & \text{si la marca de la carta no coincide con tu característica real} \end{cases}$$

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

**Modelos
cualitativos**

Modelos
cuantitativos

Paquete RReg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Warner

La respuesta aleatoria es dada por

$$r_i = \frac{z_i - (1 - p)}{2p - 1}$$

donde

- z es la variable observada
- y p es la proporción de cartas marcadas con la característica sensible

Y la varianza estimada es

$$\hat{V}_R(r_i) = r_i(r_i - 1)$$

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Warner

Un estimador insesgado para la población total de la característica sensible es

$$\hat{Y}_R = \sum_{i \in s} \frac{r_i}{\pi_i}$$

donde

- π_i es la probabilidad de inclusión de primer orden (probabilidad de que el elemento i esté en la muestra).

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Warner

La varianza de este estimador viene dada por

$$V(\hat{Y}_R) = \sum_{i \in U} \frac{V_R(r_i)}{\pi_i} + V_{HT}(r)$$

donde

- $V_R(r_i)$ es la varianza de r_i bajo el mecanismo de aleatorización y
- $V_{HT}(r)$ es la varianza del diseño del estimador de Horvitz Thompson del valor r_i .

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Warner

Esta varianza es estimada por

$$\widehat{V}(\widehat{Y}_R) = \sum_{i \in s} \frac{\widehat{V}_R(r_i)}{\pi_i} + \widehat{V}(r)$$

donde

- $\widehat{V}_R(r_i)$ varía con el mecanismo RR y
- la estimación de la varianza del diseño $\widehat{V}(r)$, se obtiene utilizando el método de Deville (Deville, 1993).

Paquete RRTCS

Función Warner

Calcula la estimación de la respuesta aleatoria, su varianza y su intervalo de confianza a través del modelo de Warner. La función también puede devolver la variable trasformada.

`Warner(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)`

- `z`: vector de la variable observada; su tamaño es igual a n (el tamaño de la muestra)
- `p`: proporción de cartas marcadas con la característica sensible
- `pi`: vector de probabilidades de inclusión de primer orden
- `type`: el tipo de estimador: “total” o “mean”
- `cl`: nivel de confianza
- `N`: tamaño de la población. Por defecto es `NULL`
- `pij`: matriz de probabilidades de inclusión de segundo orden. Por defecto es `NULL`

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Warner

- Nuestro objetivo es estimar la prevalencia sobre el abuso de alcohol
- De una población de 802 estudiantes extraemos una muestra, mediante muestreo aleatorio simple sin reemplazamiento, de tamaño 125.
- La técnica de respuesta aleatoria utilizada es el modelo de Warner (Warner, 1965) con parámetro $p = 0.7$
- La pregunta sensible es: ¿Durante el último mes, has bebido más de 5 cervezas/vinos seguidos?

Ejemplo variables cualitativas. Warner

Almacenamos en un data frame

- ID: el identificador del estudiante
- z: la respuesta a la pregunta sensible
- Pi: probabilidades de inclusión de primer orden

	ID	z	Pi
1	2204	0	0.1558603
2	238	0	0.1558603
3	318	0	0.1558603
4	450	1	0.1558603
5	5038	1	0.1558603
6	2010	1	0.1558603
7	186	1	0.1558603
8	1040	0	0.1558603
9	904	0	0.1558603
10	775	0	0.1558603

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

**Modelos
cualitativos**

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Warner

```
N=802
```

```
data(WarnerData)
```

```
dat=with(WarnerData,data.frame(z,Pi))
```

```
p=0.7
```

```
cl=0.95
```

```
Warner(dat$z,p,dat$Pi,"total",cl)
```

```
*help Warner, WarnerData
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Warner

Call:

```
Warner(z = dat$z, p = p, pi = dat$Pi,  
       type = "total", cl = cl)
```

Qualitative model

Warner model for the total estimator

Parameters: $p=0.7$

Estimation: 361

Variance: 7883.337

Confidence interval (95%)

Lower bound: 186.8784

Upper bound: 534.9216

Paquete RRTCS

Horvitz

- En el modelo Horvitz, el dispositivo de respuesta aleatoria presenta a la persona muestreada i una caja que contiene un gran número de cartas idénticas,
- con una proporción p , ($0 < p < 1$) marcadas *¿Consumes drogas?* y el resto marcadas *¿Naciste el mes de abril?* (un atributo inocuo cuya proporción poblacional α es conocida, $\frac{1}{12}$).
- La respuesta solicitada denotada por z_i toma el valor

$$z_i = \begin{cases} y_i & \text{si la carta está marcada "¿Consumes drogas?"} \\ & \text{y de verdad "Consumes drogas", o} \\ & \text{si la carta está marcada "¿Naciste el mes de abril?"} \\ & \text{y de verdad "Naciste el mes de abril"} \\ 0 & \text{si la marca de la carta no coincide con tu característica real} \end{cases}$$

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Horvitz

- Nuestro objetivo es estimar la prevalencia de los alumnos de la Universidad de Granada sobre una serie de preguntas sensibles explicadas posteriormente.
- De una población de 10777 estudiantes extraemos una muestra, mediante muestreo aleatorio simple sin reemplazamiento, de tamaño 710.
- La técnica de respuesta aleatoria utilizada es el modelo de Horvitz (Horvitz et al, 1967, Greenberg et al., 1969) con parámetro $p = 0.5$

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Diseño de la encuesta

En este modelo es necesario un cuestionario adicional con las siguientes características:

- Las preguntas no deben de ser sensibles.
- No deben estar relacionadas con el tema de la pregunta “sensible”.
- Para cada una de las preguntas que miden variables cualitativas, las respuestas deben tener una probabilidad conocida.

Paquete RRTCS

Diseño de la encuesta

CARA	¿Has copiado alguna vez en un examen?	SÍ	NO
CRUZ	¿Naciste el mes de julio?		
CARA	¿Te has peleado con algún profesor?	SÍ	NO
CRUZ	¿Tu DNI termina en número 2?		
CARA	¿Has sufrido acoso?	SÍ	NO
CRUZ	¿Naciste del 1 al 20 del mes?		
CARA	¿Has acosado alguna vez a alguien?	SÍ	NO
CRUZ	¿Tu DNI termina en número 5?		
CARA	¿Has consumido drogas en las instalaciones de la universidad?	SÍ	NO
CRUZ	¿Naciste del 15 al 25 del mes?		
CARA	¿Has mantenido relaciones sexuales en las instalaciones de la universidad?	SÍ	NO
CRUZ	¿Naciste el mes de abril?		

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Diseño de la encuesta

Para la estimación de resultados, será necesario tener en cuenta las probabilidades del cuestionario de preguntas no sensibles:

	Pregunta	Probabilidad de respuesta
1	¿Naciste el mes de julio?	1/12
2	¿Tu DNI termina en número 2?	1/10
3	¿Naciste del 1 al 20 del mes?	20/30
4	¿Tu DNI termina en número 5?	1/10
5	¿Naciste del 15 al 25 del mes?	10/30
6	¿Naciste el mes de abril?	1/12

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

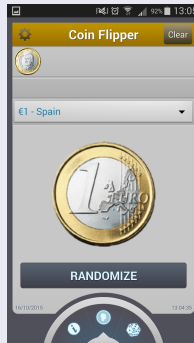
Descripción de la técnica del modelo de Horvitz

- En el desarrollo de la técnica de respuesta aleatorizada es necesario utilizar un proceso aleatorio que nos ayude a aleatorizar las respuestas.
 - El procedimiento elegido es lanzar una moneda, un lado cara y otro cruz, que va a dar lugar a un juego fácil de ejecutar.
- La anterior forma de distribución nos permite conocer fácilmente la probabilidad de que cualquier lado sea escogido al azar.

Paquete RRTCS

Descripción de la técnica del modelo de Horvitz

Con el objetivo de hacer más atractivo el mecanismo de aleatorización utilizamos la app “Randomizers” en la opción “Coin Flipper”, previamente instalada por el estudiante en su teléfono. La aplicación es muy fácil de utilizar, sólo hay que tocar el botón “Randomize” y se obtiene una cara de la moneda.



Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Descripción de la técnica del modelo de Horvitz

- Cuando se aplica la encuesta a cada estudiante de la muestra, se le pide que instale la aplicación o saque una moneda.
- Después se le pide que la lance al azar.
 - Si el lado que obtiene es cara, las preguntas que tendrá que contestar serán las sensibles.
 - Si el lado que obtiene es cruz, las preguntas que tendrá que contestar serán las no sensibles.
- De esta forma conocemos la probabilidad de que nos conteste a las preguntas sensibles, $1/2$, y como consecuencia, la probabilidad de que nos conteste a las preguntas no sensibles es de $1/2$.
- Con esto garantizamos la total aleatoriedad de respuesta.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción de la técnica del modelo de Horvitz

- Con la finalidad de darle más confianza al encuestado, se le entregó una cartilla de instrucciones:
 - Paso 1: Lanza una moneda y no le muestres a NADIE el lado obtenido
 - Paso 2: La pregunta que debes contestar en cada juego, depende del lado obtenido en la moneda:
 - Si es cara, contesta a la pregunta CARA
 - Si es cruz, contesta a la pregunta CRUZ

Este procedimiento lo debes repetir para cada juego

- El encuestado únicamente tuvo que poner en la hoja de preguntas un aspa en la respuesta que elegía y después de haber terminado de llenar los espacios con las respuestas, depositaba la ficha de la encuesta en una urna.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción de la técnica del modelo de Horvitz

Almacenamos en un data frame las respuestas a las 6 preguntas sensibles (`HorvitzDataRealSurvey.rda`)

Resultados

Los resultados que se obtuvieron después de la aplicación de la encuesta son sumamente importantes, ya que a partir de éstos podemos hacer inferencias, comparaciones y contrastes.

Estimación por medio del modelo de Horvitz

Se asignan los datos conocidos de las probabilidades de las preguntas no sensibles, así como el número de estudiantes que respondieron afirmativamente a cada una de las preguntas.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Resultados

Tamaño de la muestra=710 estudiantes, Tamaño de la población=10777

Pregunta	Respuestas sí en la muestra	p	1-p	Probabilidad de la característica no sensitiva
1	328	0.5	0.5	1/12
2	180	0.5	0.5	1/10
3	300	0.5	0.5	20/30
4	101	0.5	0.5	1/10
5	164	0.5	0.5	10/30
6	53	0.5	0.5	1/12

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Horvitz

```
N=10777
```

```
n=710
```

```
data(HorvitzDataRealSurvey)
```

```
p=0.5
```

```
pi=rep(n/N,n)
```

```
cl=0.95
```

```
alpha=c(1/12,1/10,20/30,1/10,10/30,1/12)
```

```
Horvitz(HorvitzDataRealSurvey$...,p,alpha[...],  
        pi,"mean",cl,N)
```

```
*help Horvitz, HorvitzDataRealSurvey
```


Paquete RRTCS

Resultados

El cálculo de la proporción y la estimación de la característica sensible se presenta en la siguiente tabla, para la muestra total utilizando la fórmula del modelo de Horvitz.

Al tener las proporciones estimadas a partir de la muestra, se procede a calcular las varianzas de los estimadores de las preguntas sensibles.

Pregunta	Proporción	Estimación de la característica sensible	Varianza
1	0.4619718	0.8406103	0.001389716
2	0.2535211	0.4070423	0.001045196
3	0.3943662	0.1220657	0.001337415
4	0.1140845	0.128169	0.000559785
5	0.2309859	0.1286385	0.000991658
6	0.07464789	0.06596244	0.000383954

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzStCl

- Nuestro objetivo es estimar la prevalencia sobre la infidelidad
- De una población de 1500 estudiantes extraemos una muestra, mediante muestreo estratificado (facultad) por conglomerados (grupos), de tamaño 365.
- La técnica de respuesta aleatoria utilizada es el modelo de Horvitz (Horvitz et al., 1967, Greenberg et al., 1969) con parámetro $p = 0.6$
- La pregunta sensible es: ¿Has sido infiel alguna vez?
- La pregunta no relacionada es: ¿Tu DNI termina en número impar? con una probabilidad $\alpha = 0.5$

Ejemplo variables cualitativas. HorvitzStCl

Almacenamos en un data frame

- ID: el identificador del estudiante
- ST: el identificador del estrato
- CL: el identificador del cluster
- z: la respuesta a la pregunta sensible
- Pi: probabilidades de inclusión de primer orden

	ID	ST	CL	z	Pi
1	28	1	2	1	0.28
2	22	1	2	1	0.28
3	30	1	2	1	0.28
4	29	1	2	0	0.28
5	27	1	2	1	0.28
6	23	1	2	0	0.28
7	26	1	2	1	0.28
8	24	1	2	1	0.28
9	21	1	2	1	0.28
10	25	1	2	0	0.28

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzStCl

```
N=1500
data(HorvitzDataStCl)
dat=with(HorvitzDataStCl, data.frame(ST,CL,z,Pi))
p=0.6
alpha=0.5
cl=0.95
Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)
```

*help Horvitz, HorvitzDataStCl, ResamplingVariance

Paquete RRTCS

Ejemplo variables cualitativas. HorvitzStCl

Call:

```
Horvitz(z = dat$z, p = p, alpha = alpha,  
pi = dat$Pi, type = "mean", cl = cl, N = N)
```

Qualitative model

Horvitz model for the mean estimator

Parameters: $p=0.6$; $\alpha=0.5$

Estimation: 0.3939394

Variance: 0.001688151

Confidence interval (95%)

Lower bound: 0.3134101

Upper bound: 0.4744687

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Función ResamplingVariance

También podemos calcular la varianza mediante métodos de remuestreo, mediante la función ResamplingVariance

```
ResamplingVariance(output,pi,type=c("total","mean"),  
option=1,N=NULL,pij=NULL,str=NULL,clu=NULL,srswr=FALSE)
```

- output: salida del método cualitativo o cuantitativo dependiendo de la variable de interés
- pi: vector de probabilidades de inclusión de primer orden. Por defecto es NULL
- type: tipo del estimador: “total” o “mean”
- option: método usado para calcular la varianza (1: Jackknife, 2: Escobar-Berger, 3: Campbell-Berger-Skinner). Por defecto es 1

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Función ResamplingVariance

- N: tamaño de la población
- pij: matriz de probabilidades de inclusión de segundo orden. Esta matriz es necesaria para las opciones de Escobar-Berger y Campbell-Berger-Skinner. Por defecto es NULL
- str: identificador del estrato. Este vector es necesario para la opción Jackknife. Por defecto es NULL
- clu: identificador del cluster. Este vector es necesario para la opción Jackknife. Por defecto es NULL
- srswr: variable que indica si el muestreo es con reemplazamiento. Por defecto es NULL

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzStCl

```
out=Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)
```

- Método Jackknife

```
ResamplingVariance(out,dat$Pi,"mean",N=N,  
str=dat$ST,clu=dat$CL)
```

```
[1] 0.001885475
```

- Método Campbell-Berger-Skinner

```
ResamplingVariance(out,dat$Pi,"mean",3,N,  
samplingVarEst::Pkl.Hajek.s(dat$Pi))
```

```
[1] 0.001443204
```


Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

SoberanisCruz

- El modelo SoberanisCruz considera la introducción de una variable inocua correlacionada con la variable sensible.
- Esta variable no afecta a la sensibilidad individual, y mantiene la fiabilidad.
- El procedimiento de muestreo es el mismo del modelo Horvitz
- Ejemplo:
 - La pregunta sensible es: ¿Excedes los límites de velocidad a menudo?
 - La pregunta no relacionada es: Es tu coche de gama medio/alta? con probabilidad $\alpha = 0.5$

Paquete RRTCS

ForcedResponse

- En el esquema de respuesta forzada, a la persona muestreada i se le ofrece una caja con cartas:
- algunas están marcadas con *Sí* con una proporción p_1 , algunas marcadas con *No* con una proporción p_2 y el resto están marcadas con *¿Consumes drogas?*, en la proporción restante $p_3 = 1 - p_1 - p_2$, donde $0 < p_1, p_2 < 1$, $p_1 \neq p_2$, $p_1 + p_2 < 1$.
- Se le pide a la persona que extraiga aleatoriamente una de ellas, observe la marca de la carta y responda

$$z_i = \begin{cases} 1 & \text{si la carta es de tipo "Sí"} \\ 0 & \text{si la carta es de tipo "No"} \\ y_i & \text{si la carta es de tipo "¿Consumes drogas?"} \end{cases}$$

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. ForcedResponse

- Tenemos un conjunto de datos de una población simulada, de tamaño 10000, mediante una distribución binomial con probabilidad 0.5.
- Extraemos una muestra, mediante muestreo aleatorio simple con reemplazamiento, de tamaño 1000.
- La técnica de respuesta aleatoria utilizada es el modelo ForcedResponse (Boruch, 1972) con parámetros $p_1 = 0.2$ y $p_2 = 0.2$

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. ForcedResponse

```
set.seed(123)
N=10000
y=rbinom(N,1,0.5)
n=1000
ys=sample(y,n,replace=TRUE)
pi=rep(1-(1-(1/N))^n,n)
p1=0.2
p2=0.2
tb=100
box=c(rep(1,p1*tb),rep(2,p2*tb),rep(3,(1-p1-p2)*tb))
extra=sample(box,n,replace=TRUE)
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. ForcedResponse

```
z=vector()
z[extra==1]=1
z[extra==2]=0
z[extra==3]=ys[extra==3]
FRD=data.frame(z,pi,row.names=NULL)
names(FRD)=c("z","Pi")
cl=0.95
ForcedResponse(FRD$z,p1,p2,FRD$Pi,"mean",cl,N)
```

*help ForcedResponse

Paquete RRTCS

Ejemplo variables cualitativas. ForcedResponse

Call:

```
ForcedResponse(z = FRD$z, p1 = p1, p2 = p2,  
pi = FRD$Pi, type = "mean", cl = cl, N = N)
```

Qualitative model

Forced Response model for the total estimator

Parameters: p1=0.2; p2=0.2

Estimation: 0.5464073

Variance: 0.0007407925

Confidence interval (95%)

Lower bound: 0.493062

Upper bound: 0.5997526

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list
endorse

Bibliografía

HorvitzUB

- En el modelo Horvitz, cuando la proporción de la población α es desconocida, se toman dos muestras independientes.
- Se llenan dos cajas con un gran número de cartas similares excepto que
 - en la primera caja una proporción de ellas p_1 están marcadas *¿Consumes drogas?* y la proporción complementaria $(1 - p_1)$ lleva la marca *¿Realizas deporte?*,
 - mientras que en la segunda caja estas proporciones son p_2 y $1 - p_2$, manteniendo p_2 distinto de p_1 .
- Se elige una muestra y se pide a cada persona muestreada que extraiga una carta aleatoriamente de la primera caja y lo repita independientemente con la segunda caja.

Paquete RRTCS

HorvitzUB

- En el primer caso, debe dar una respuesta aleatoria, como

$$I_i = \begin{cases} 1 & \text{si la carta est marcada “¿Consumes drogas?”} \\ & \text{y de verdad “Consumes drogas” o} \\ & \text{si la carta est marcada “¿Realizas deporte?”} \\ & \text{y de verdad “Realizas deporte”} \\ 0 & \text{si la marca de la carta de la primera caja} \\ & \text{no coincide con tu caracterstica real} \end{cases}$$

- y en el segundo caso, dar una respuesta aleatoria como

$$J_i = \begin{cases} 1 & \text{si la carta est marcada “¿Consumes drogas?”} \\ & \text{y de verdad “Consumes drogas” o} \\ & \text{si la carta est marcada “¿Realizas deporte?”} \\ & \text{y de verdad “Realizas deporte”} \\ 0 & \text{si la marca de la carta de la segunda caja} \\ & \text{no coincide con tu caracterstica real} \end{cases}$$

Beatriz
Cobo
Rodrguez
Mara del
Mar Rueda
Garca

Introduccin

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografa

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzUB

- Nuestro objetivo es estimar la prevalencia sobre el abuso de drogas
- De una población de 802 estudiantes extraemos una muestra, mediante muestreo por conglomerados con probabilidades proporcionales al tamaño, de tamaño 188.
- La técnica de respuesta aleatoria utilizada es el modelo de HorvitzUB (Chaudhuri, 2011) con parámetros $p_1 = 0.6$ y $p_2 = 0.7$)
- La pregunta sensible en ambas cajas es: ¿Has consumido drogas alguna vez?
- La pregunta no relacionada es: ¿Prefieres el fútbol al tenis?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzUB

```
N=802
data(HorvitzUBData)
dat=with(HorvitzUBData,data.frame(I,J,Pi))
p1=0.6
p2=0.7
cl=0.95
HorvitzUB(dat$I,dat$J,p1,p2,dat$Pi,"mean",cl,N)
```

```
*help HorvitzUB, HorvitzUBData
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. HorvitzUB

Call:

```
HorvitzUB(I = dat$I, J = dat$J, p1 = p1, p2 = p2,  
pi = dat$Pi, type = "mean", cl = cl, N = N)
```

Qualitative model

Horvitz unknown B model for the mean estimator

Parameters: p1=0.6; p2=0.7

Estimation: 0.1213169

Variance: 0.00758909

Confidence interval (95%)

Lower bound: -0.0494261

Upper bound: 0.2920599

Paquete RRTCS

Kuk

- En la técnica de respuesta aleatoria de Kuk, a la persona muestreada i se le ofrecen dos cajas.
- Cada caja contiene cartas que son idénticas a excepción del color, ya sea rojo o blanco, en un número suficientemente grande,
 - con proporciones p_1 y $1 - p_1$ en la primera y
 - p_2 y $1 - p_2$, en la segunda ($p_1 \neq p_2$).
- Se le pide a la persona muestreada usar
 - la primera caja, si “*Consume drogas*” y
 - la segunda caja, si “*No consume drogas*”y hacer k extracciones independientes de cartas, con reemplazamiento cada vez.
- Debe responder $z_i = f_i$, siendo f_i el número de veces que ha extraído una carta roja.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Kuk

- Nuestro objetivo es estimar la prevalencia sobre el exceso de actividad sexual
- De una población de 802 estudiantes extraemos una muestra, mediante muestreo aleatorio simple sin reemplazamiento, de tamaño 200.
- La técnica de respuesta aleatoria utilizada es el modelo de Kuk (Kuk, 1990) con parámetros $p_1 = 0.6$, $p_2 = 0.2$ y $k = 25$
- La pregunta sensible es: ¿Realizas actividades sexuales de forma excesiva?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. Kuk

```
N=802
```

```
data(KukData)
```

```
dat=with(KukData,data.frame(z,Pi))
```

```
p1=0.6
```

```
p2=0.2
```

```
k=25
```

```
cl=0.95
```

```
Kuk(dat$z,p1,p2,k,dat$Pi,"mean",cl,N)
```

```
*help Kuk, KukData
```

Paquete RRTCS

Ejemplo variables cualitativas. Kuk

Call:

```
Kuk(z = dat$z, p1 = p1, p2 = p2, k = k, pi = dat$Pi,  
    type = "mean", cl = cl, N = N)
```

Qualitative model

Kuk model for the mean estimator

Parameters: p1=0.6; p2=0.2; k=25

Estimation: 0.1335

Variance: 0.0006541851

Confidence interval (95%)

Lower bound: 0.08336992

Upper bound: 0.1836301

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

MangatSingh

- En el modelo Mangat-Singh, a la persona muestreada se le ofrecen dos cajas de cartas.
- En la primera caja una proporción conocida t , ($0 < t < 1$) de cartas está marcada con *¿Consumes drogas?* y el resto están marcadas con *Randomized Response*.
- Se extrae una carta, se observa y se devuelve a la caja.
 - Si la carta está marcada *¿Consumes drogas?*, entonces el encuestado debería responder “Sí”, si *Consume drogas*, de lo contrario responde “No”.
 - Si la carta extraída está marcada *Randomized Response*, entonces el encuestado debe utilizar la segunda caja y extraer una carta de ella.
 - Esta segunda caja contiene una proporción p , ($0 < p < 1$), ($p \neq 0.5$) de cartas marcadas *¿Consumes drogas?* y el resto marcadas *¿No consumes drogas?*.
Si la carta extraída de la segunda caja coincide con su característica, debe responder “Sí”, de lo contrario responde “No”.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

MangatSingh

- La respuesta aleatoria de la persona etiquetada i se asume que es:

$$z_i = \begin{cases} y_i & \text{si la carta marcada "¿Consumes drogas?"} \\ & \text{es extraída de la primera caja} \\ I_i & \text{si la carta está marcada "RR"} \\ & \text{es extraída de la primera caja} \end{cases}$$

$$I_i = \begin{cases} 1 & \text{si la carta está marcada "¿Consumes drogas?"} \\ & \text{y de verdad "Consumes drogas"} \\ & \text{si la carta está marcada "¿No consumes drogas?"} \\ & \text{y de verdad "No consumes drogas"} \\ 0 & \text{si la marca de la carta de la segunda caja} \\ & \text{no coincide con tu característica real} \end{cases}$$

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. MangatSingh

- Nuestro objetivo es estimar la prevalencia sobre el abuso de cannabis
- De una población de 802 estudiantes extraemos una muestra, mediante muestreo muestreo estratificado por año académico, de tamaño 240.
- La técnica de respuesta aleatoria utilizada es el modelo de Mangat-Singh (Mangat y Singh, 1990) con parámetros $p = 0.7$ y $t = 0.55$
- La pregunta sensible es: ¿Has consumido cannabis alguna vez?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cualitativas. MangatSingh

```
N=802
data(MangatSinghData)
dat=with(MangatSinghData,data.frame(z,Pi))
p=0.7
t=0.55
cl=0.95
MangatSingh(dat$z,p,t,dat$Pi,"mean",cl,N)
```

```
*help MangatSingh, MangatSinghData
```

Paquete RRTCS

Ejemplo variables cualitativas. MangatSingh

Call:

```
MangatSingh(z = dat$z, p = p, t = t, pi = dat$Pi,  
             type = "mean", cl = cl, N = N)
```

Qualitative model

Mangat and Singh model for the mean estimator

Parameters: $p=0.7$; $t=0.55$

Estimation: 0.5004562

Variance: 0.001650115

Confidence interval (95%)

Lower bound: 0.4208393

Upper bound: 0.5800731

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete list

Paquete
endorse

Bibliografía

Ejemplo

Nuestro objetivo es estimar el número de porros que fuman a la semana los estudiantes.

Para ello:

- extraemos una muestra mediante un mecanismo de muestreo
- aplicamos las distintas técnicas cuantitativas de respuesta aleatoria disponibles en el paquete RRTCS

Paquete RRTCS

BarLev

- A una persona de la muestra etiquetada i se le ofrece una caja con un número considerable de cartas idénticas
- con una proporción $p(0 < p < 1; p \neq 0.5)$ de ellas marcadas *¿Cuántos porros fumas a la semana?* y el resto marcadas *Randomized Response*.
- Se le pide a la persona que extraiga aleatoriamente una de ellas, observe la marca de la carta y de una respuesta

$$z_i = \begin{cases} y_i & \text{si la carta extraída es "¿Cuántos porros fumas a la semana?"} \\ y_i S & \text{si la carta extraída es "Randomized Response"} \end{cases}$$

donde S es una variable de aleatorización, cuya media μ y desviación estándar σ son conocidas.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función BarLev

Calcula la estimación de la respuesta aleatoria, su varianza, y su intervalo de confianza a través del modelo BarLev. La función también puede devolver la variable transformada.

```
BarLev(z,p,mu,sigma,pi,type=c("total","mean"),cl,  
N=NULL,pij=NULL)
```

- z : vector de la variable observada; su longitud es igual a n (el tamaño de la muestra)
- p : probabilidad de respuesta directa

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función BarLev

- mu: media de la variable aleatoria S
- sigma: desviación estándar de la variable aleatoria S
- pi: vector de probabilidades de inclusión de primer orden
- type: tipo del estimador: “total” o “mean”
- cl: nivel de confianza
- N: tamaño de la población. Por defecto es NULL
- pij: matriz de probabilidades de inclusión de segundo orden. Por defecto es NULL

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. BarLev

- Nuestro objetivo es estimar los ingresos de las empresas.
- De una población de 2396 empresas extraemos una muestra, mediante muestreo estratificado con probabilidades proporcionales al tamaño de la compañía, de tamaño 370.
- La técnica de respuesta aleatoria utilizada es el modelo de BarLev (Bar-Lev et al, 2004) con parámetros $p = 0.6$ y variable de aleatorización $S = \exp(1)$
- La pregunta sensible es: ¿Cuáles fueron los ingresos de la compañía en el año fiscal anterior?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. BarLev

En caso de tener una encuesta directa podemos transformarla a encuesta por respuesta aleatoria

```
#para utilizar el conjunto de datos
```

```
library(TeachingSampling)
```

```
data(Lucy)
```

```
attach(Lucy)
```

```
N=2396
```

```
mean(Income)*N
```

```
[1] 1035217
```

```
#vector de tamaño de muestra en cada estrato
```

```
nh=c(70,100,200)
```

```
n=sum(nh)
```

```
set.seed(123)
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. BarLev

```
#vector que identifica la pertenencia a los  
#estratos de cada unidad de la población  
S=Level  
#vector de información auxiliar para cada unidad  
#de la población  
x=Employees  
#extrae una muestra aleatoria simple con probabilidad  
#proporcional al tamaño sin reemplazamiento  
#de tamaño nh en el estrato h de tamaño Nh  
res=S.STpiPS(S,x,nh)  
id=res[,1] #unidad seleccionada  
pi=res[,2] #probabilidad de inclusión  
sdata=Lucy[id,]
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. BarLev

```
p=0.6
tb=100
box=c(rep("A",p*tb),rep("B",(1-p)*tb))
extra=sample(box,n,replace=TRUE)
S=rexp(n)
lambda=1
mu=1/lambda
sigma=1/lambda^2
z=sdata$Income
z[extra=="B"]=z[extra=="B"]*S[extra=="B"]
BarLevData=data.frame(id,sdata$Level,z,pi,row.names=NULL)
names(BarLevData)=c("ID","ST","z","Pi")
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. BarLev

```
data(BarLevData)
dat=with(BarLevData,data.frame(z,Pi))
p=0.6
mu=1
sigma=1
cl=0.95
BarLev(dat$z,p,mu,sigma,dat$Pi,"total",cl)
```

```
*help BarLev, BarLevData
```

Paquete RRTCS

Ejemplo variables cuantitativas. BarLev

Call:

```
BarLev(z = dat$z, p = p, mu = mu, sigma = sigma,  
       pi = dat$Pi, type = "total", cl = cl)
```

Quantitative model

Bar Lev model for the total estimator

Parameters: $p=0.6$; $\mu=1$; $\sigma=1$

Estimation: 1019372

Variance: 3313677571

Confidence interval (95%)

Lower bound: 906547.7

Upper bound: 1132197

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

EichhornHayre

La respuesta aleatoria dada por la persona etiquetada i es

$$z_i = y_i S$$

donde S es una variable de aleatorización cuya distribución es conocida.

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. EichhornHayre

- Nuestro objetivo es estimar los ingresos familiares.
- Extraemos una muestra, mediante muestreo estratificado por propiedad de la vivienda, de tamaño 150.
- La técnica de respuesta aleatoria utilizada es el modelo de Eichhorn y Hayre (Eichhorn y Hayre, 1983) con variable de aleatorización $S = F(20, 20)$
- La pregunta sensible es: ¿Cuáles son los ingresos familiares anuales?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. EichhornHayre

```
data(EichhornHayreData)
dat=with(EichhornHayreData,data.frame(ST,z,Pi))
mu=1.111111
sigma=0.5414886
cl=0.95
#This line returns a warning showing why the variance
estimation is not possible.
#See ResamplingVariance for several alternatives.
EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
```

*help EichhornHayre, EichhornHayreData, ResamplingVariance

Paquete RRTCS

Ejemplo variables cuantitativas. EichhornHayre

Call:

```
EichhornHayre(z = dat$z, mu = mu, sigma = sigma,  
pi = dat$Pi, type = "mean", cl = cl)
```

Quantitative model

Eichhorn and Hayre model for the mean estimator

Parameters: mu=1.1; sigma=0.54

Estimation: 38048.79

Variance:

Confidence interval (95%)

Lower bound:

Upper bound:

Warning message:

In Estimator(out1, pi, type, cl, N, pij) :

To calculate the estimated variance is needed
or the size of the population or the second-order
inclusion probabilities matrix.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. EichhornHayre

```
out=EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
ResamplingVariance(out,dat$Pi,"mean",1,str=dat$ST)
[1] 2179498
```

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Eriksson

- A una persona de la muestra etiquetada i se le ofrece una caja con un número considerable de cartas idénticas
- con una proporción $p(0 < p < 1; p \neq 0.5)$ de ellas marcadas *¿Cuántos porros fumas a la semana?* y el resto marcadas m_1, \dots, m_j con probabilidades q_1, \dots, q_j verificando $q_1 + \dots + q_j = 1 - p$.
- Se le pide a la persona que extraiga aleatoriamente una de ellas y observe la marca de la carta.

La respuesta aleatoria dada por la persona etiquetada i es:

$$z_i = \begin{cases} y_i & \text{si la carta extraída es "¿Cuántos porros fumas a la semana?"} \\ S & \text{si la carta extraída es "m_1, \dots, m_j"} \end{cases}$$

donde S es una variable uniforme discreta

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. Eriksson

- Nuestro objetivo es estimar el comportamiento de copia en los exámenes.
- De una población de 53376 estudiantes extraemos una muestra, mediante muestreo estratificado por facultad universitaria con afijación uniforme, de tamaño 102.
- La técnica de respuesta aleatoria utilizada es el modelo de Eriksson (Eriksson, 1973) con parámetro $p = 0.5$ y S una variable uniforme discreta en los puntos (0, 1, 3, 5, 8).
- La pregunta sensible es: ¿Cuántas veces copiaste en los exámenes el año pasado?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Descripción de la técnica del modelo de Eriksson

Selecciona una carta de la baraja sin que nadie te vea.

- Si sale 1,2,3,4 o 5 escribe la respuesta a la pregunta sensible en el apartado para contestar.
- Si sale otro, escribe el número que aparece junto a la imagen correspondiente.

Paquete RRTCS

Descripción de la técnica del modelo de Eriksson

Con el objetivo de hacer más atractivo el mecanismo de aleatorización utilizamos la app “Baraja Española”, previamente instalada por el estudiante en su teléfono.

La aplicación es muy fácil de utilizar, sólo hay que tocar la pantalla y se obtiene una carta de la bajara.



Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Diseño de la encuesta

- ¿Cuánto dinero gastaste en alcohol el último fin de semana?



- ¿Cuántos porros fumas a la semana?



- ¿Cuántas veces copiaste en exámenes el año pasado?



Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. Eriksson

¿Cuántas veces copiaste en exámenes el año pasado?

```
N=53376
```

```
data(ErikssonData)
```

```
dat=with(ErikssonData,data.frame(z,Pi))
```

```
p=0.5
```

```
mu=mean(c(0,1,3,5,8))
```

```
sigma=sqrt(4/5*var(c(0,1,3,5,8)))
```

```
cl=0.95
```

```
Eriksson(dat$z,p,mu,sigma,dat$Pi,"mean",cl,N)
```

```
*help Eriksson, ErikssonData
```

Paquete RRTCS

Ejemplo variables cuantitativas. Eriksson

Call:

```
Eriksson(z = dat$z, p = p, mu = mu, sigma = sigma,  
         pi = dat$Pi, type = "mean", cl = cl, N = N)
```

Quantitative model

Eriksson model for the mean estimator

Parameters: $p=0.5$; $\mu=3.4$; $\sigma=2.9$

Estimation: 4.44307

Variance: 1.549224

Confidence interval (95%)

Lower bound: 2.003546

Upper bound: 6.882595

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

DianaPerri1

- A una persona de la muestra etiquetada i se le ofrece una caja con un número considerable de cartas idénticas
- con una proporción $p(0 < p < 1; p \neq 0.5)$ de ellas marcadas *¿Cuántos porros fumas a la semana?* y el resto marcadas *Randomized Response*.
- Se le pide a la persona que extraiga aleatoriamente una de ellas, observe la marca de la carta y de una respuesta

$$z_i = \begin{cases} y_i & \text{si la carta extraída es “¿Cuántos porros fumas a la semana?”} \\ W(y_i + U) & \text{si la carta extraída es “Randomized Response”} \end{cases}$$

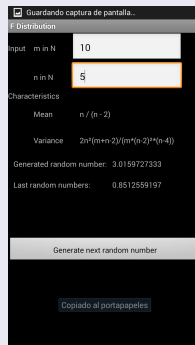
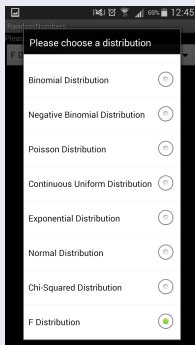
donde W, U son variables de aleatorización cuya distribución es conocida.

Paquete RRTCS

Diseño de la encuesta

Para calcular los valores aleatorios de las distribuciones utilizamos la app “RandomNumbers”.

Hay que elegir la distribución, asignarle los parámetros y pulsar “Generate next random number”.



Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list
endorse

Bibliografía

Ejemplo variables cuantitativas. DianaPerri1

- Nuestro objetivo es estimar el fraude de impuestos.
- De una población de 417 individuos extraemos una muestra, mediante muestreo aleatorio simple sin reemplazamiento, de tamaño 150.
- La técnica de respuesta aleatoria utilizada es el modelo de Diana y Perri 1 (Diana y Perri, 2010) con parámetros $p = 0.6$, $W = F(10, 5)$ y $U = F(5, 5)$.
- La pregunta sensible es: ¿Qué cantidad de subsidio agrícola declaras en la declaración de la renta?

Paquete RRTCS

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo variables cuantitativas. DianaPerri1

```
N=417
```

```
data(DianaPerri1Data)
```

```
dat=with(DianaPerri1Data,data.frame(z,Pi))
```

```
p=0.6
```

```
mu=c(5/3,5/3)
```

```
cl=0.95
```

```
DianaPerri1(dat$z,p,mu,dat$Pi,"mean",cl,N,"srswor")
```

```
*help DianaPerri1, DianaPerri1Data
```

Paquete RRTCS

Ejemplo variables cuantitativas. DianaPerri1

Call:

```
DianaPerri1(z = dat$z, p = p, mu = mu, pi = dat$Pi,  
type = "mean", cl = cl, N = N, method = "srswor")
```

Quantitative model

Diana and Perri model for the mean estimator

Parameters: $p=0.6$; $\mu_1=1.7$; $\mu_2=1.7$

Estimation: 6411.94

Variance: 1137327

Confidence interval (95%)

Lower bound: 4321.726

Upper bound: 8502.153

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción

El paquete RRreg: Correlation and Regression Analyses for Randomized Response Data (Heck y Moshagen, 2014):

- Realiza métodos univariantes y multivariantes para analizar diseños de encuestas con respuesta aleatoria (RR).
- Se pueden utilizar las variables RR para correlaciones, como variable dependiente en una regresión logística y como predictor en una regresión lineal.
- Para fines de simulación y bootstrap, los datos RR se pueden generar de acuerdo a varios modelos.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Descripción

Las principales funciones proporcionan las siguientes funcionalidades:

- RRuni: análisis univariante simple, es decir, las estimaciones de prevalencia
- RRcor: correlaciones bivariantes incluyendo variables de RR
- RRlog: regresión logística con una variable RR como criterio
- RRlin: regresión lineal con un criterio continuo, no RR, incluyendo variables RR como predictores.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción

Además, se pueden utilizar dos funciones para generar datos de estudios robustos, estimaciones bootstrap, y propósitos de pruebas

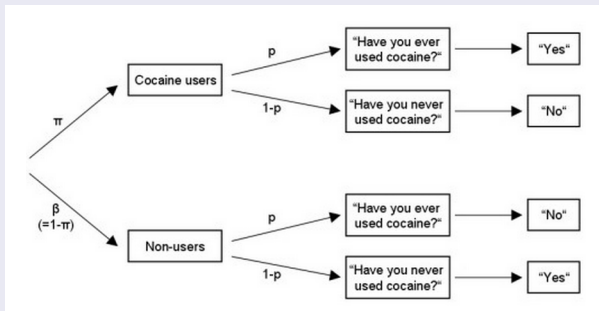
- **RRgen**: genera un único marco de datos que incluye los estados verdaderos, las respuestas RR, y la pertenencia a los grupos
- **RRsimu**: simulación Monte Carlo para probar **RRuni**, **RRcor** y **RRlog**, ya sea para un RR y una variable continua no RR o para dos variables RR

Paquete RRreg

Ejemplo análisis univariante. Warner

$$\begin{cases} \text{¿Alguna vez ha consumido cocaína?}, & p \\ \text{¿Usted nunca ha consumido cocaína?}, & 1 - p \end{cases}$$

Como mecanismo de aleatorización, se pueden utilizar dados o una moneda con probabilidades conocidas.



Paquete RRreg

Función RRgen

Genera datos de acuerdo a un modelo de respuesta aleatorio específico.

```
RRgen(n,pi.true,model,p,complyRates=c(1, 1),  
sysBias=c(0,0),groupRatio=0.5,Kukrep=1,trueState=NULL)
```

- n: tamaño de la muestra de los datos generados
- pi.true: proporción verdadera de la población
- model: especifica el modelo de respuesta aleatorio,
 - “Warner”, “UQTknown”, “UQTunknown”, “Mangat”, “Kuk”, “FR”, “Crosswise”, “CDM”, “CDMsym”, “SLD”, “mix.norm”, “mix.exp”.
- p: probabilidad de aleatorización (dependiendo del modelo)
- complyRates: vector con dos valores dando las proporciones de portadores y no portadores que se adhieren a las instrucciones, respectivamente

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Función RRgen

- *sysBias*: probabilidad de responder ‘sí’ (codificado como 1) en caso de no conformidad de los portadores y no portadores del atributo sensible, respectivamente.
 - Si *sysBias* = c(0,0), los portadores y no portadores dan sistemáticamente la respuesta no sensible ‘no’.
 - Si *sysBias* = c(0,0.5) los portadores siempre responden ‘no’, mientras que los no portadores seleccionan al azar una categoría de respuesta.
 - Si *sysBias* = c(0.5,0.5) podría ser la mejor opción para *Kuk* y *Crosswise*.
 - Nota: Para el modelo *FR* con m -categorías, *sysBias* se puede dar como un vector de probabilidades para las categorías de 0 a $(m - 1)$

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Función RRgen

- groupRatio: proporción de participantes en el grupo 1. Sólo se requiere para modelos de dos grupos.
- Kukrep: número de repeticiones del procedimiento de Kuk.
- trueState: vector opcional que contiene verdaderos estados de los participantes que se asignarán al azar de acuerdo con el procedimiento definido.
 - 1 para los portadores y
 - 0 para los no portadores de atributos sensibles;
 - Nota: para FR valores entre 1 y el número de categorías de respuesta
 - si se especifican, n y $pi.true$ estos son ignorados

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis univariante. Warner

Simulamos datos para 1000 participantes, con una proporción del 30 % de los usuarios cocainomanos, y una probabilidad de aleatorización de $p = 0.2$.

```
library(RRreg)
data.W=RRgen(n=1000,pi.true=.3,model="Warner",p=.2)
head(data.W)
```

##	true	comply	response
## 1	0	1	1
## 2	1	1	1
## 3	1	1	0
## 4	0	1	0
## 5	0	1	1
## 6	0	1	1

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis univariante. Warner

Ahora que tenemos un conjuntos de datos, podemos estimar la prevalencia de consumidores de cocaína a través de RRuni.

Función RRuni

Analiza un vector de datos *response* con un modelo de RR especificado con probabilidad de aleatorización conocida p

`RRuni(response,data,model,p,group=NULL,Mlest=TRUE)`

- *response*: ya sea un vector de respuestas que contiene 0='no' y 1='sí' o el nombre de la variable respuesta en *data*.
 - En el método de juegos de cartas de Kuk, la variable respuesta observada da el número de cartas rojas.
 - Para el modelo de *Forced Response*, los valores de respuesta son números enteros de 0 a $(m - 1)$, donde m es el número de categorías de respuesta
- *data*: *data.frame* opcional que contiene la variable respuesta

Paquete RRreg

Función RRuni

- model: define el modelo de RR.
 - Modelos disponibles: “Warner”, “UQTknown”, “UQTunknown”, “Mangat”, “Kuk”, “FR”, “Crosswise”, “CDM”, “CDMsym”, “SLD”, “mix.norm”, “mix.exp”, “mix.unknown”.
- p: probabilidad de aleatorización definida como una única probabilidad para
 - Warner: probabilidad de obtener la pregunta sensible.
 - Mangat: probabilidad de los no portadores para responder verazamente.
 - Crosswise: prevalencia de respuestas ‘sí’ para la pregunta no relacionada (la categoría de respuesta se codifica como $1=[\text{‘no-no’ o ‘sí-sí’}]$; $0=[\text{‘sí-no’ o ‘no-sí’}]$)

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRuni

y como un vector 2-valuado de probabilidades para

- Kuk: probabilidad de cartas rojas en primer y segundo grupo, respectivamente (rojo=1, negro=0)
- Unrelated Question (UQTknown): probabilidad para responder a la pregunta sensible y prevalencia conocida de respuestas 'sí' para la pregunta no relacionada
- Unrelated Question (UQTunknown): probabilidad para responder a la pregunta sensible en el grupo 1 y 2, respectivamente.
- Cheating Detection (CDM): probabilidad para ser solicitado a decir que 'sí' en el grupo 1 y 2, respectivamente.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Función RRuni

- Symmetric CDM: vector 4-valuado. Probabilidad para ser solicitado a decir que 'sí'/'no' en el grupo 1 y 'sí'/'no' en el grupo 2
- Stochastic Lie Detector (SLD): probabilidad para los no portadores para responder con 0='no' en el grupo 1 y 2, respectivamente
- Forced Response model (FR): vector m-valuado (m =número de categorías de respuesta) con las probabilidades de ser solicitado para seleccionar las categorías de respuesta 0,1,..., $m-1$, respectivamente (requiere $\sum(p) < 1$)

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRuni

Para modelos RR continuos

- `mix.norm`: vector 3-valuado. Probabilidad para responder a la pregunta sensible y media y desviación estándar de la distribución normal enmascarada de la pregunta no relacionada
- `mix.exp`: vector 2-valuado. Probabilidad para responder a la pregunta sensible y media de la distribución exponencial enmascarada de la pregunta no relacionada
- `mix.unknown`: vector 2-valuado. Probabilidad de responder a la pregunta sensible en el grupo 1 y 2, respectivamente.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRuni

- **group**: un vector de grupos de la misma longitud que *response* conteniendo los valores 1 ó 2,
 - sólo se requiere para los modelos de dos grupos, que especifican diferentes probabilidades de aleatorización para los dos grupos.
 - Si se proporciona un `data.frame` *data*, la variable *group* se busca dentro de él.
- **MLest**: si es TRUE, las estimaciones mínimos cuadráticas de pi fuera de $[0,1]$ se corrigen para obtener estimaciones máximo verosímiles.

Paquete RRreg

Ejemplo análisis univariante. Warner

```
warner=RRuni(response=response,data=data.W,  
model="Warner",p=.2)  
summary(warner)  
## Call:  
## Warner Model with p = 0.2  
## Sample size: 1000  
##  
##           Estimate   StdErr      z  Pr(>|z|)  
##      pi 0.318333 0.025719 12.378 < 2.2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05  
## '.' 0.1 ' ' 1
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

Estudiamos ahora la correlación entre “Consumir cocaína” y “Número de veces que vas a la discoteca por semana”.

Generamos una variable continua, no una variable RR.

Según nuestra simulación, los encuestados que “Consumen cocaína” tienen puntuaciones más altas en “Número de veces que vas a la discoteca por semana”

```
data.W$cov[data.W$true==1]=rnorm(sum(data.W$true==1),3,1)
data.W$cov[data.W$true==0]=rnorm(sum(data.W$true==0),1,1)
```

Ahora, se puede estimar la correlación bivariada entre la variable de RR de Warner dicotómica y la covariable continua.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Función RRcor

RRcor calcula las correlaciones de Pearson bivariantes de variables medidas con o sin RR.

```
RRcor(x,y=NULL,models,p.list,group=NULL,bs.n=0,  
bs.type=c("se.n","se.p","pval"),nCPU=1)
```

- *x*: un vector numérico, matriz o data frame
- *y*: *NULL* (por defecto) o un vector, matriz o data frame con dimensiones compatibles a *x*.
- *models*: un vector definiendo que diseño de RR se utiliza para cada variable.
 - Debe estar en el mismo orden en el que aparece en las variables *x* e *y* (por columnas).
 - Modelos discretos disponibles: *Warner*, *Kuk*, *FR*, *Mangat*, *UQTknown*, *UQTunknown*, *Crosswise*, *SLD* y *directo* (es decir, no hay diseño de respuesta aleatoria).
 - Modelos continuos disponibles: *mix.norm*, *mix.exp*.

Paquete RRreg

Función RRcor

- `p.list`: *list* que contiene las probabilidades de aleatorización de los modelos RR definidos en *models*.
 - O bien, todas las variables *direct* en *models* pueden ser excluidas en *p.list*;
 - o, si se especifican, las probabilidades de aleatorización *p* se ignoran para variables *direct*.
- `group`: una matriz que define la pertenencia al grupo de cada participante (valores 1 y 2) para todos los modelos de grupos múltiples.
 - Si sólo uno de estos modelos se incluye en *models*, se puede utilizar un vector.
 - Para más de un modelo, cada columna debe contener una variable de agrupación.
- `bs.n`: número de muestras utilizadas para obtener errores estándar bootstrap.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRcor

- `bs.type`: para obtener errores estándar bootstrap,
 - utilice “*se.p*” para bootstrap paramétrico.
El bootstrap paramétrico se basa en la suposición de que la variable continua se distribuye normalmente dentro de los grupos definidos por el verdadero estado de la variable de RR.
 - y/o “*se.n*” para noparamétrico.
 - Utilice “*pval*” para obtener los p-valores del bootstrap paramétrico (suponiendo una correlación verdadera de cero).
 - Tenga en cuenta que *bs.n* tiene que ser mayor que 0.
 - Para diseños de respuesta forzada politómica (FR), la variable RR se asume que tiene distancias equidistantes entre las categorías.
- `nCPU`: número de CPUs utilizados para el bootstrap

Paquete RRreg

Ejemplo análisis multivariante. Warner

```
RRcor(x=data.W$response,y=data.W$cov,  
models=c("Warner","direct"),p.list=list(.2))  
## Randomized response variables:  
##   Variable          RRmodel  p  
## 1 data.W$response Warner    0.2  
## 2 data.W$cov         direct  
##  
## Sample size N = 1000  
##  
## Estimated correlation matrix:  
##               data.W$response data.W$cov  
## data.W$response      1.000000  0.689248  
## data.W$cov           0.689248  1.000000
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

Podemos ejecutar una regresión logística para predecir la probabilidad del “Consumo de cocaína” a través del “Número de veces que vas a la discoteca por semana”.

Paquete RRreg

Función RRlog

Una variable dicotómica, medida por un método de respuesta aleatoria, sirve como variable dependiente utilizando uno o más predictores continuos y/o categóricos

```
RRlog(formula,data,model,p,group,LR.test=TRUE,fit.n=1,  
EM.max=1000,optim.max=500,...)
```

- formula: especificar el modelo de regresión
- data: *data.frame*, en el que se pueden encontrar las variables
- model: modelos RR disponibles:
 - “Warner”, “UQTknown”, “UQTunknown”, “Mangat”, “Kuk”, “FR”, “Crosswise”, “CDM”, “CDMsym”, “SLD”.
- p: probabilidad/probabilidades de aleatorización.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRlog

- **group**: vector especificando la pertenencia al grupo.
 - Puede ser omitido para diseños RR de un grupo único.
 - Para diseños RR de dos grupos, utilice 1 y 2 para indicar la pertenencia al grupo, igualando las probabilidades de aleatorización respectivas $p[1]$ y $p[2]$.
 - Si un diseño RR y una pregunta directa (DQ) fueran utilizadas en el estudio, los índices de grupo se establecen en 0 (DQ) y 1 (RR; 1 o 2 para diseños RR de dos grupos).

Función RRlog

- Esto puede ser utilizado para probar, si el diseño de RR da lugar a una estimación de prevalencia diferente mediante la inclusión de una variable dummy para el formato de pregunta (RR vs DQ) como predictor.

Si el coeficiente de regresión correspondiente es significativo, las estimaciones de prevalencia varían entre RR y DQ.

- Del mismo modo, se pueden probar las hipótesis de interacción
 - por ejemplo, la correlación entre un atributo sensible y un predictor se encuentra sólo utilizando un diseño de RR pero no con un diseño de respuesta directa.
- Las hipótesis de este tipo pueden ser probadas mediante la inclusión de la interacción de la variable dummy-RR-DQ y el predictor en *formula*
 - por ejemplo, $RR \sim dummy * predictor$

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRlog

- LR.test: coeficientes de regresión del test mediante el test de razón de verosimilitudes, es decir, ajustar el modelo en varias ocasiones mientras se excluye un parámetro a la vez
- fit.n: número de repeticiones de ajuste utilizando valores iniciales aleatorios para evitar parámetros máximos locales *fit.bound*.
 - El modelo es ajustado repetidamente hasta que las estimaciones del parámetro absoluto están por debajo de *fit.bound* o se alcanza el número máximo de repeticiones de ajuste.
 - De este modo, se aumenta la estabilidad de las estimaciones.
 - *fit.bound* debe aumentarse si las estimaciones de los parámetros extremos son las esperadas.

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función RRlog

- EM.max: número máximo de iteraciones del algoritmo EM.
 - Si $EM.max = 0$, el algoritmo EM se omite
- optim.max: número máximo de iteraciones en cada ejecución de *optim*

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

En *formula*, la variable RR de Warner se define como criterio en el lado izquierdo, mientras que la covariable continua se utiliza como predictor a la derecha.

```
log1=RRlog(formula=response~cov,data=data.W,  
model="Warner",p=.2)  
summary(log1)
```

Paquete RRreg

Ejemplo análisis multivariante. Warner

```
## Call:
## RRlog.formula(formula = response ~ cov, data = data.W,
##               model = "Warner", p = 0.2)
##
## RR Model:
## Warner with p = 0.2
##
## Model fit:
##      n      logLik
## 1000 -578.0839
##
##              Estimate   StdErr Wald test Pr(>Chi2,df=1)
## (Intercept) -4.79562    0.74027   41.96715      0.00000
## cov          1.98309    0.31985   38.43956      0.00000
##
##              deltaG2    Pr(>deltaG2)
## (Intercept)  227.02      < 2.2e-16 ***
## cov          175.81      < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

Si queremos quitar la constante del modelo, debemos escribir

```
formula=response~cov-1
```

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

También podemos utilizar la variable RR de Warner “Consumir cocaína” como predictor en una regresión lineal para explicar el “Número de veces que vas a la discoteca por semana”

Paquete RRreg

Función RRlin

Regresión lineal para un criterio continuo, utilizando variables de respuesta aleatoria (RR) como predictores

```
RRlin(formula,data,models,p.list,group=NULL,Kukrep=1,  
bs.n=0,nCPU=1,maxit=1000,fit.n=5,pibeta=0.05)
```

- `formula`: un criterio continuo es predicho por una o más variables RR categoricas definidas por *models*.
 - Si el número de predictores excede el número definido por el vector *models*, los predictores restantes se tratan como variables no aleatorias.
 - Interacciones incluyendo algunas variables RR no pueden ser incluidas.
- `data`: un data frame opcional, lista o entorno, que contiene las variables en el modelo

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Función RRlin

- **models**: vector de caracteres especificando el modelo RR en el orden de aparición en *formula*.
 - Modelos disponibles: “Warner”, “UQTknown”, “UQTunknown”, “Mangat”, “Kuk”, “FR”, “Crosswise”, “CDM”, “CDMsym”, “SLD”, “custom”
- **p.list**: lista de probabilidades de aleatorización para los modelos RR en el mismo orden en el que se especifica en *models*.
 - Tenga en cuenta, que las probabilidades de aleatorización *p* deben ser proporcionadas en una estructura *list*.
- **group**: vector o matriz especificando la pertenencia al grupo mediante los índices 1 y 2.
 - Solo para los modelos RR multigrupos.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Función RRlin

- Kukrep: define el número de repeticiones en el método de juego de cartas de Kuk
- bs.n: número de muestras utilizadas para el bootstrap no paramétrico
- nCPU: número de núcleos utilizados para el bootstrap
- maxit: número máximo de iteraciones en la rutina de optimización
- fit.n: número de ejecuciones de ajuste con valores de comienzo aleatorios
- pibeta: razón aproximada de probabilidades p_i para los pesos de regresión beta.
 - Puede ser utilizado para la aceleración y estimación ML afinada.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete RRreg

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

```
lin1=RRlin(formula=cov~response,data=data.W,  
models="Warner",p.list=.2,fit.n=1)  
summary(lin1)
```

Paquete RRreg

Ejemplo análisis multivariante. Warner

```
## Call:
## RRlin(formula = cov ~ response, data = data.W,
##       models = "Warner", p.list = 0.2, fit.n = 1)
##
## Randomized response variables:
##   Variable Model   p
## 1 response Warner 0.2
##
## Coefficients (beta):
##               Estimate StdErr Wald test Pr(>Chi2,df=1)
## (Intercept)  1.04245 0.04900   452.56      0.01465 *
## response      1.94442 0.08573   514.36      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error (sigma): 0.975; N=1000
##
## Prevalence estimates for combinations of RR responses:
##   Estimate StdErr
## 0  0.67581 0.0212
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo análisis multivariante. Warner

Además de la salida de la regresión, la función *summary* también ofrece ‘Estimaciones de prevalencia para combinaciones de respuestas RR’.

- En nuestro ejemplo, como únicamente tenemos una variable de respuesta aleatoria, nos indica un 0, y muestra la estimación de la prevalencia para el subgrupo que no tiene el atributo sensible.
- En el caso de tener dos atributos sensibles:
 - La combinación 0:0, indica que el subgrupo no tiene ninguno de los dos atributos sensibles y muestra la estimación de la prevalencia correspondiente.
 - Del mismo modo, la combinación 0:1 indica el subgrupo que tiene sólo el segundo atributo sensible y así sucesivamente.
 - Tenga en cuenta que la última combinación 1:1 no se proporciona y se puede calcular sumando el resto de las estimaciones de prevalencia y restando esta cantidad a 1.

Descripción

El paquete rr: Statistical Methods for the Randomized Response Technique (Blair, Zhou e Imai, 2015)

- Implementa métodos desarrollados por Blair, Imai y Zhou (2015)
- Permite realizar análisis de regresión multivariante para variables sensibles bajo cuatro diseños de respuesta aleatoria estándar,
 - mirrored question (pregunta duplicada), forced response, disguised response (respuesta disfrazada), y unrelated question

Paquete rr

Descripción

- Realiza la regresión con la respuesta aleatoria como resultado y la regresión logística con el elemento de respuesta aleatoria como predictor.
- Genera probabilidades predichas de responder afirmativamente a la pregunta sensible para cada encuestado.
- Permite a los usuarios utilizar el elemento sensible como un predictor en una salida de regresión bajo el diseño *forced response*.
- Se implementa el análisis de potencia para el diseño de elementos de respuesta aleatoria para ayudar a mejorar el diseño de la investigación.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Descripción

El paquete consta de funciones para llevar a cabo diversos análisis:

- `rrreg`: permite a los usuarios realizar análisis de regresión multivariante de datos de la técnica de respuesta aleatoria. El método implementado por esta función es la estimación máximo verosímil (ML) para el algoritmo de esperanza-maximización (EM)
- `rrreg.predictor`: permite a los usuarios realizar análisis de regresión multivariante con el elemento de respuesta aleatoria como un predictor de un resultado separado de interés. Lo hace mediante el modelado conjuntamente del elemento de respuesta aleatoria como resultado y predictor para una salida adicional dado el mismo conjunto de covariables.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Descripción

- `predict.rrreg`: permite a los usuarios generar probabilidades predichas para el elemento de respuesta aleatorio dado un objeto de clase “rrreg” de la función *rrreg* anterior
- `predict.rrreg.predictor`: permite a los usuarios generar probabilidades predichas para las variables de resultado adicionales con el elemento de respuesta aleatoria como covariable dado un objeto de clase “rrreg.predictor” de la función *rrreg.predictor*
- Además, el paquete incluye la función *power.rr.test*. Esta función permite a los usuarios realizar un análisis de potencia para diseños de encuestas por respuesta aleatoria, para los cuatro diseños previamente descritos.

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo Modelo Respuesta Forzada

Queremos ajustar una regresión multivariante para predecir los ciudadanos que “Tienen contacto social directo a grupos armados” a través de una serie de predictores

Paquete rr

Función rrreg

rrreg se utiliza para llevar a cabo un análisis de regresión multivariante de datos de la encuesta utilizando métodos de respuesta aleatorios.

```
rrreg(formula,p,p0,p1,q,design,data,start=NULL,  
maxIter=10000,verbose=FALSE,optim=FALSE,  
em.converge=10^(-8),glmMaxIter=10000,  
solve.tolerance=.Machine$double.eps)
```

- *formula*: un objeto de clase “formula”: una descripción simbólica del modelo que se ajusta

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función rrreg

- p :
 - la probabilidad de recibir la pregunta sensible (Mirrored Question Design, Unrelated Question Design);
 - la probabilidad de responder verazmente (Forced Response Design);
 - la probabilidad de seleccionar una carta roja del montón de ‘sí’ (Disguised Response Design).
 - Para diseños “mirrored” y “disguised”, p no puede ser igual a 0.5.
- p_0 : la probabilidad de ‘no’ forzado (Forced Response Design)
- p_1 : la probabilidad de ‘sí’ forzado (Forced Response Design)

Paquete rr

Función rrreg

- q: la probabilidad de responder ‘sí’ a la pregunta no relacionada que se supone que es independiente de las covariables (Unrelated Question Design)
- design: uno de los cuatro diseños estándar:
 - “forced-known”, “mirrored”, “disguised”, o “unrelated-known”
- data: un data frame que contiene las variables en el modelo
- start: valores iniciales opcionales de estimaciones de los coeficientes para el algoritmo de esperanza-maximización (EM)
- maxIter: número máximo de iteraciones para el algoritmo de esperanza-maximización. Por defecto es 10000.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función rrreg

- `verbose`: un valor lógico que indica si el diagnóstico del modelo cuenta el número de iteraciones EM que se imprime. Por defecto es `FALSE`.
- `optim`: un valor lógico que indica si utilizar el metodo “BFGS” quasi-Newton para calcular la matriz de varianza-covarianza y errores estándar. Por defecto es `FALSE`
- `em.converge`: un valor que especifica el grado satisfactorio de convergencia bajo el algoritmo EM. Por defecto es $10^{(-8)}$
- `glmMaxIter`: un valor que especifica el número máximo de iteraciones para ejecutar el algoritmo EM. Por defecto es 10000
- `solve.tolerance`: cuando se calculan los errores estándar, esta opción especifica la tolerancia de resolver la operación de inversión de la matriz

Paquete rr

Ejemplo Modelo Respuesta Forzada

```
library(rr)
data(nigeria)
set.seed(1)
p=2/3 #probabilidad de responder honestamente en el
#diseño de respuesta forzada
p1=1/6 #probabilidad de un 'Sí' forzado
p0=1/6 #probabilidad de un 'No' forzado

rr.q1.reg.obj=rrreg(rr.q1~cov.asset.index+cov.married+
I(cov.age/10)+I((cov.age/10)^2)+cov.education+cov.female,
data=nigeria,p=p,p1=p1,p0=p0,design="forced-known")
summary(rr.q1.reg.obj)
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Ejemplo Modelo Respuesta Forzada

Randomized Response Technique Regression

```
Call: rrreg(formula = rr.q1 ~ cov.asset.index + cov.married +  
I(cov.age/10) + I((cov.age/10)^2) + cov.education + cov.female,  
p = p, p0 = p0, p1 = p1, design = "forced-known", data = nigeria)
```

	Est.	S.E.
(Intercept)	-0.34017	0.50856
cov.asset.index	0.07896	0.04136
cov.married	-0.26743	0.25451
I(cov.age/10)	-0.35283	0.26423
I((cov.age/10)^2)	0.04099	0.02603
cov.education	-0.00691	0.04558
cov.female	-0.55439	0.16244

Randomized response forced design with $p = 0.67$, $p0 = 0.17$,
and $p1 = 0.17$.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo Modelo Respuesta Forzada

Queremos ajustar una regresión multivariante para predecir los ciudadanos que “se unen a un grupo cívico” a través de una serie de predictores en el que se incluye la variable de respuesta aleatoria “Tener contacto directo a grupos armados”.

Y además ajustar una regresión multivariante para predecir los ciudadanos que “Tienen contacto directo a grupos armados” con las variables predictoras anteriores.

Paquete rr

Función rrreg.predictor

rrreg.predictor se utiliza para modelar conjuntamente el elemento de respuesta aleatoria como el resultado y el predictor para un resultado adicional dado un conjunto de covariables

```
rrreg.predictor(formula,p,p0,p1,q,design,data,rr.item,  
model.outcome="logistic",fit.sens="bayesglm",  
fit.outcome="bayesglm",bstart=NULL,tstart=NULL,  
parstart=TRUE,maxIter=10000,verbose=FALSE,optim=FALSE,  
em.converge=10^(-4),glmMaxIter=20000,estconv=TRUE,  
solve.tolerance=.Machine$double.eps)
```

- *formula*: un objeto de clase “formula”: una descripción simbólica del modelo que se ajusta con el elemento de respuesta aleatoria como una de las covariables.

Función rrreg.predictor

- p:
 - la probabilidad de recibir la pregunta sensible (Mirrored Question Design, Unrelated Question Design);
 - la probabilidad de responder verazmente (Forced Response Design);
 - la probabilidad de seleccionar una carta roja del montón de 'sí' (Disguised Response Design).
- p0: la probabilidad de 'no' forzado (Forced Response Design)
- p1: la probabilidad de 'sí' forzado (Forced Response Design)
- q: la probabilidad de responder 'sí' a la pregunta no relacionada que se supone que es independiente de las covariables (Unrelated Question Design)

Paquete rr

Función rrreg.predictor

- design: uno de los cuatro diseños estándar:
 - “forced-known”, “mirrored”, “disguised”, o “unrelated-known”
- data: un data frame que contiene las variables en el modelo.
 - Las observaciones con valores perdidos son prudentemente eliminadas
- rr.item: una cadena que contiene el nombre de la variable de respuesta aleatoria en el data frame
- model.outcome: actualmente la función sólo permite la regresión logística, es decir, la variable resultado debe ser binaria.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Función rrreg.predictor

- `fit.sens`: indicador de si utilizar el modelado lineal generalizado Bayesiano (`bayesglm`) en el paso de maximización para el algoritmo de esperanza-maximización (EM) para generar coeficientes para el elemento de respuesta aleatoria como resultado.
 - Por defecto es “`bayesglm`”;
 - en otro caso la entrada es “`glm`”
- `fit.outcome`: indicador de si utilizar el modelado lineal generalizado bayesiano (`bayesglm`) en el paso de maximización para el algoritmo EM para generar coeficientes para la variable de resultado dada en la formula con el elemento de respuesta aleatoria como una covariable.
 - Por defecto es “`bayesglm`”;
 - en otro caso la entrada es “`glm`”

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función rrreg.predictor

- `bstart`: valores iniciales opcionales de estimaciones de los coeficientes para el elemento de respuesta aleatoria como resultado para el algoritmo EM
- `tstart`: valores iniciales opcionales de estimaciones de los coeficientes de la variable de resultado dada en la fórmula para el algoritmo EM
- `parstart`: opción para utilizar la función `rrreg` para generar valores iniciales de las estimaciones de los coeficientes para el elemento de respuesta aleatoria como resultado para el algoritmo EM.
 - Por defecto es `TRUE`,
 - pero si las estimaciones iniciales son introducidas por el usuario en `bstart`, esta opción se anula.

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Función rrreg.predictor

- `maxIter`: número máximo de iteraciones para el algoritmo de esperanza-maximización. Por defecto es 10000.
- `verbose`: un valor lógico que indica si el diagnóstico del modelo cuenta el número de iteraciones EM que se imprime. Por defecto es FALSE.
- `optim`: un valor lógico que indica si utilizar el método “BFGS” quasi-Newton para calcular la matriz de varianza-covarianza y errores estándar. Por defecto es FALSE
- `em.converge`: un valor que especifica el grado satisfactorio de convergencia bajo el algoritmo EM. Por defecto es $10^{(-4)}$

Paquete rr

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función rrreg.predictor

- `glmMaxIter`: un valor que especifica el número máximo de iteraciones para ejecutar el algoritmo EM. Por defecto es 20000
- `estconv`: opción a la convergencia base en el valor absoluto de la diferencia entre los coeficientes siguientes generados a través del algoritmo EM en lugar de las posteriores log-verosimilitudes. Por defecto es TRUE
- `solve.tolerance`: cuando se calculan los errores estándar, esta opción especifica la tolerancia de resolver la operación de inversión de la matriz

Paquete rr

Ejemplo Modelo Respuesta Forzada

```
data(nigeria)
set.seed(44)
p=2/3 #probabilidad de responder honestamente en el
#diseño de respuesta forzada
p1=1/6 #probabilidad de un 'Sí' forzado
p0=1/6 #probabilidad de un 'No' forzado

rr.q1.pred.obj=rrreg.predictor(civic~cov.asset.index+
cov.married+I(cov.age/10)+I((cov.age/10)^2)+
cov.education+cov.female+rr.q1,rr.item="rr.q1",
parstart=FALSE,data=nigeria,optim=TRUE,
p=p,p1=p1,p0=p0,design="forced-known")
summary(rr.q1.pred.obj)
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete rr

Ejemplo Modelo Respuesta Forzada

Randomized Response as a Regression Predictor

```
Call: rrreg.predictor(formula = civic ~ cov.asset.index + cov.married +  
  I(cov.age/10) + I((cov.age/10)^2) + cov.education + cov.female +  
  rr.q1, p = p, p0 = p0, p1 = p1, design = "forced-known",  
  data = nigeria, rr.item = "rr.q1", parstart = FALSE,  
  optim = TRUE)
```

	est.t	se.t	est.b	se.b
(Intercept)	-2.83808	0.31178	-0.51932	0.46869
cov.asset.index	0.11211	0.02271	0.08683	0.04069
cov.married	0.31915	0.13342	-0.32189	0.24012
I(cov.age/10)	0.94090	0.15851	-0.25157	0.24722
I((cov.age/10)^2)	-0.10259	0.01676	0.03168	0.02542
cov.education	0.08606	0.02513	-0.01536	0.04442
cov.female	-0.02532	0.08852	-0.56205	0.16242
rr.q1	0.38404	0.15978	NA	NA

Randomized response forced design with $p = 0.67$, $p_0 = 0.17$, and $p_1 = 0.17$.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Resumen comparativo

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

RRTCS(1)	RRreg(2)	rr(3)
Christofides	-	-
-	CDM(cheating detection model)	-
-	CDMsym	-
-	Crosswise	-
Devore	-	-
ForcedResponse	FR \neq 1, 3	forced response
Horvitz	UQTknown	unrelated question
HorvitzUB	UQTunknown	-
Kuk	Kuk	disguised response \neq 1, 2
Mangat	Mangat \neq 1	-
MangatUB	-	-
MangatSingh	-	-
MangatSinghSingh	-	-
MangatSinghSinghUB	-	-
SinghJoarder	-	-
-	SLD(stochastic lie detector)	-
SoberanisCruz	-	-
Warner	Warner	mirrored question
BarLev	-	-
ChaudhuriChristofides	-	-
DianaPerri1	-	-
DianaPerri2	-	-
EichhornHayre	-	-
Eriksson	-	-
Saha	-	-

Paquete list

Descripción

Paquete list: Statistical Methods for the Item Count Technique and List Experiment (Blair, Imai, Park y Coppock, 2010)

- Implementa una alternativa a los métodos de respuesta aleatorizada: la técnica de conteo de items (Chaudhuri and Christofides 2007)
- Esta metodología de encuestas también se conoce como lista de experimentos o técnica de conteo incomparable.
- Bajo un diseño estándar, se pregunta a los encuestados acerca de un conjunto de acciones o puntos de vista a la vez en lugar de uno sensible de forma aislada.
- Para evaluar la prevalencia de actitudes y comportamientos sensibles, el investigador aleatoriza si el elemento sensible de interés se añade a la lista de elementos de control.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Descripción

- El paquete implementa los métodos desarrollados por Imai (2011) y Blair y Imai (2012), Blair, Imai, y Lyall (2013), Imai, Park, y Greene (2014), y Aronow, Coppock, Crawford, y Green (2015).
- Permite a los investigadores llevar a cabo análisis estadísticos multivariantes de datos de encuestas con la lista de experimentos.
- Una implementación MCMC bayesiana de la regresión para diseños de lista de experimentos con elementos sensibles estándar y múltiples y una configuración de efectos aleatorios
- un modelo de regresión jerárquico MCMC bayesiano con hasta tres grupos jerárquicos

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción

- lista de experimentos combinada
- modelo de regresión de experimento de respaldo
- un modelo conjunto de la lista de experimentos que permite el análisis de la lista de experimentos como predictor en modelos de regresión
- un método para combinar la lista de experimentos con preguntas directas
- Además implementa la prueba estadística que se diseña para detectar ciertos fracasos de la lista de experimentos
- y una prueba de placebo para la lista de experimentos utilizando datos de preguntas directas.
- El paquete también incluye dos conjuntos de datos que contienen las aplicaciones de la técnica de conteo de items en dos encuestas nacionales.

Paquete list

Ejemplo

Supongamos que tenemos 1500 individuos, 1000 de los cuales participan, pero 500 de ellos se niegan a participar si se les pregunta directamente.

```
# Definimos los tipos de sujetos
# Verdaderamente responden "Sí" a la pregunta directa
VRS=500
# Falsamente responden "No" a la pregunta directa
FRN=500
# Verdaderamente responder "No" a la pregunta directa
VRN=500
```

```
tipo=rep(c("VRS","FRN","VRN"),times=c(VRS,FRN,VRN))
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo. Pregunta directa

Ahora supongamos que fuéramos a hacer la pregunta, ¿Quiere usted participar?

```
D=ifelse(tipo=="VRS",1,0)
direct.est=mean(D)
direct.est
## [1] 0.3333333
```

- La verdadera proporción participantes es $1000/1500 = 0.67$.
- Sin embargo, la pregunta directa está sesgada por la deseabilidad social, nuestra estimación de pregunta directa es 0.33

Paquete list

Lista de experimentos convencional

La lista de experimentos convencional aborda la deseabilidad social preguntando a

- un grupo control, ¿Cuántos de J comportamientos (no sensibles) tienes?
 - He estado en Irlanda
 - A menudo veo la televisión por la noche
 - Tengo más de una hermana
- un grupo tratamiento, ¿Cuántos de J+1 comportamientos tienes?, donde el comportamiento adicional es el sensible.
 - He estado en Irlanda
 - A menudo veo la televisión por la noche
 - He consumido cannabis en la universidad
 - Tengo más de una hermana

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Lista de experimentos convencional

La diferencia de las medias es una estimación de la prevalencia que está libre de sesgo de deseabilidad social. Esta estimación se basa en dos supuestos adicionales:

- sin mentirosos: requiere que los sujetos de tratamiento respondan con veracidad a la pregunta de la lista
- sin efectos del diseño: requiere que la presencia del elemento sensible no cambie las respuestas de los sujetos de tratamiento a los elementos no sensibles

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Ejemplo. Lista de experimentos convencional

```
N=length(tipo)
```

```
# Genera respuestas list de posibles resultados
```

```
# Resultados posibles de control
```

```
Y0=sample(1:4,N,replace=TRUE)
```

```
# Resultados posibles de tratamientos, es 1 más  
grande para los que verdaderamente responden Sí  
y para los que falsamente responden No
```

```
Y1=Y0+ifelse(tipo %in% c("VRS", "FRN"),1,0)
```

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete
endorse

Bibliografía

Ejemplo. Lista de experimentos convencional

```
# Realiza la asignación aleatoria
```

```
Z=rbinom(N,1,0.5)
```

```
# Revela las respuestas de list
```

```
Y=Z*Y1+(1-Z)*Y0
```

```
list.est=mean(Y[Z==1])-mean(Y[Z==0])
```

```
list.se=sqrt((var(Y[Z==1])/sum(Z)+var(Y[Z==0])/sum(1-Z)))
```

```
list.est
```

```
## [1] 0.6334232
```

```
list.se
```

```
## [1] 0.05996451
```

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr
Paquete list

Paquete
endorse

Bibliografía

Ejemplo. Lista de experimentos convencional

- La lista de experimentos convencional se acerca más a la verdad, nuestra estimación es ahora 0.63, siendo la proporción de participantes del 0.67.
- El error estándar es algo más grande, 0.06.
- La principal dificultad con el uso de experimentos list es que las estimaciones pueden ser imprecisas.

Paquete list

Lista de experimentos combinada

- El propósito del estimador combinado es aumentar la precisión mediante la combinación de interrogatorio directo con la lista de experimentos.
- La estimación combinada es un promedio ponderado de la estimación de pregunta directa y la estimación de la lista de experimentos entre los que respondieron “No” a la pregunta directa.
- Bajo dos supuestos adicionales:
 - independencia: requiere que el tratamiento no tenga efecto sobre la respuesta de la pregunta directa
 - monotonía: requiere que no haya temas “falsamente confesados” a la pregunta directa

el estimador combinado produce estimaciones más precisas que el estimador convencional

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Función combinedListDirect

Esta función implementa el estimador de lista combinada descrito en Aronow, Coppock, Crawford, y Green (2015): Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence

```
combinedListDirect(formula,data=parent.frame(),  
treat="treat",direct="direct")
```

- formula: un objeto de la clase “formula”: una descripción simbólica del modelo que se ajusta.
- Debe ser de la forma $Y \sim T + X1 + X2$, donde
 - Y es la respuesta de list,
 - T es el indicador de tratamiento, y
 - X1, X2, etc son covariables de pretatamiento.

Se recomienda que T sea una variable numérica cuyos valores son

- 0 para los sujetos control y
- 1 para los sujetos en tratamiento.

Paquete list

Función combinedListDirect

- **data**: data frame opcional, lista o entorno que contiene las variables en el modelo.
 - Si no se encuentra en *data*, las variables se toman del medio (*formula*), típicamente se llama el medio de *combined.list*.
 - Es una buena práctica incluir todas las variables utilizadas en la estimación (lista de respuesta, indicador de tratamiento, respuesta directa, y covariables de pretatamiento opcionales) en un data frame, en lugar de llamar a los datos del medio global.
- **treat**: cadena de caracteres con el nombre de la variable de tratamiento. Por defecto “treat”
- **direct**: cadena de caracteres con el nombre de la variable de respuesta directa. Por defecto “direct”.
 - La variable de respuesta directa en sí debe contener solo valores 0 y 1, donde 1 se refiere a los sujetos que respondieron “Sí” a la pregunta directa.

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RReg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Ejemplo. Lista de experimentos combinada

```
library(list)

# Incluya todos los datos en un data.frame
df=data.frame(Y,Z,D)

out.1=combinedListDirect(formula=Y~Z,data=df,
treat="Z",direct="D")

out.1

## Combined List Estimates
##
## Call: combinedListDirect(formula = Y ~ Z,
## data = df, treat = "Z", direct = "D")
##
## Prevalence estimate
##              Prevalence
## Estimate      0.65863351
## Standard Error 0.04987643
```

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Paquete list

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list
endorse

Bibliografía

Ejemplo. Lista de experimentos combinada

- La lista de experimentos combinada tiene una estimación de 0.65, siendo la proporción de participantes 0.67.
- Si comparamos los errores estándar de los dos métodos, podemos ver que el estimador combinado es más preciso que el estimador convencional.

Paquete endorse

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Descripción

Paquete endorse: R Package for Analyzing Endorsement Experiments (Shiraito y Imai, 2012)

- Implementa el modelo estadístico propuesto por Bullock, Imai, y Shapiro (2011, Análisis Político) para analizar experimentos de respaldo.
- Los experimentos de respaldo son una metodología de encuesta para inducir respuestas veraces ante preguntas sensibles.
- Esta metodología es útil cuando se mide el apoyo a los protagonistas políticos socialmente sensibles, tales como grupos militantes.
- El modelo se ajusta con el algoritmo Monte Carlo para cadenas de Markov y produce la salida que contiene la extracción para la distribución posterior.

Bibliografía



G. Blair, K. Imai and Y.Y. Zhou. *Package ‘rr’: Statistical Methods for the Randomized Response Technique*. URL = <http://cran.r-project.org/web/packages/rr/> (2015).



G. Blair, K. Imai, B. Park and A. Coppock. *Package ‘list’: Statistical Methods for the Item Count Technique and List Experiment*. URL = <http://cran.r-project.org/web/packages/list/> (2010).



B. Cobo, M. Rueda, A. Arcos. *Package ‘RRTCS’: Randomized Response Techniques for Complex Surveys*. URL = <http://cran.r-project.org/web/packages/RRTCS/> (2015).



D.W. Heck and M. Moshagen. *Package ‘RRreg’: Correlation and Regression Analyses for Randomized Response Data*. URL = <http://cran.r-project.org/web/packages/RRreg/> (2014).



Y. Shiraito and K. Imai. *Package ‘endorse’: R Package for Analyzing Endorsement Experiment*. URL = <http://cran.r-project.org/web/packages/endorse/> (2012).



S.L. Warner. *Randomized response: A survey technique for eliminating evasive answer bias*. JASA, 60, 63-69 (1965).

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Agradecimientos

Beatriz
Cobo
Rodríguez
María del
Mar Rueda
García

Introducción

Paquetes R

Paquete
RRTCS

Modelos
cualitativos

Modelos
cuantitativos

Paquete RRreg

Paquete rr

Paquete list

Paquete
endorse

Bibliografía

Este estudio ha sido parcialmente financiado por el Ministerio de Educación, Cultura y Deporte (proyecto MTM2012-35650 y programa de becas FPU, España) y por Consejería de Economía, Innovación, Ciencia y Empleo (proyecto SEJ2954, Junta de Andalucía, España).