

Causales

s = muestra teórica

s_{IN} = no elegibles.

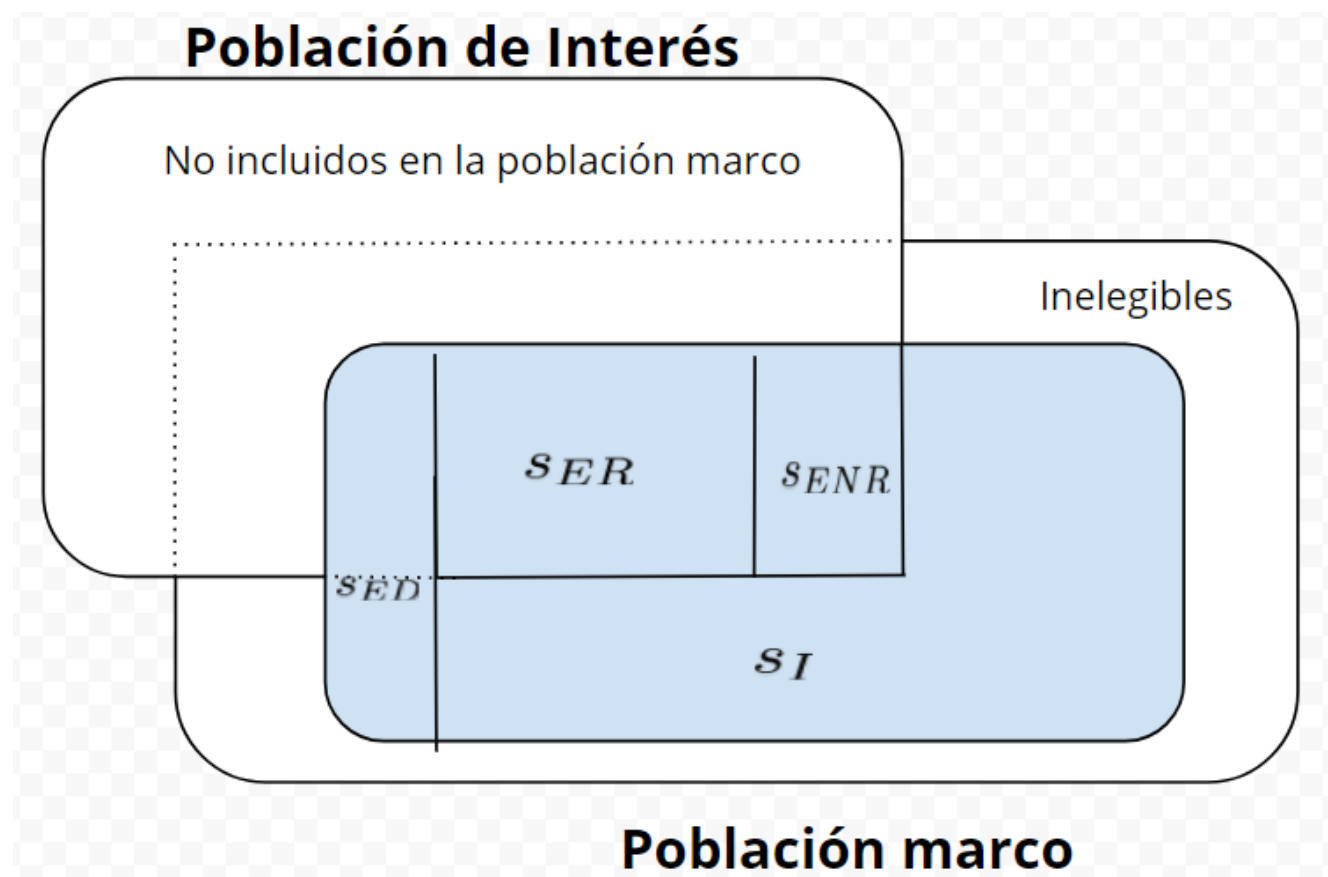
s_{ER} = elegibles respondentes

s_{ENR} = elegibles no respondentes

s_{EC} = elegibilidad conocida ($s_{IN} \cup s_{ER} \cup s_{ENR}$)

s_{ED} = elegibilidad desconocida.

Causales



Ajuste por elegibilidad desconocida

- Por variadas razones puede ocurrir que no sea posible determinar si la unidad es elegible (**e.g. dirección no especificada**).
- Si hay unidades en la muestra que no son elegibles, implica que también las hay en s_{EC}
- El mecanismo de **ajuste suele ser muy sencillo**
 - Poco se sabe de las unidades en s_{ED}
 - La no respuesta es el mayor problema

Ajuste por elegibilidad desconocida

- Un método es distribuir los pesos de las unidades incluidas en s_{DE} entre las unidades incluidas en s_{EC} .
 - Si las tasas de desconocimiento de elegibilidad varían mucho, es recomendable hacerlo a nivel de clases/grupos de ajuste
 - Las clases se crean con información del marco (generalmente geográfica).

Ajuste por elegibilidad desconocida

La Idea:

1. Se forman $b = 1, \dots, B$ clases basadas en información del marco. Las clases pueden cruzar los estratos.
2. El ajuste por desconocimiento de elegibilidad en la clase b es

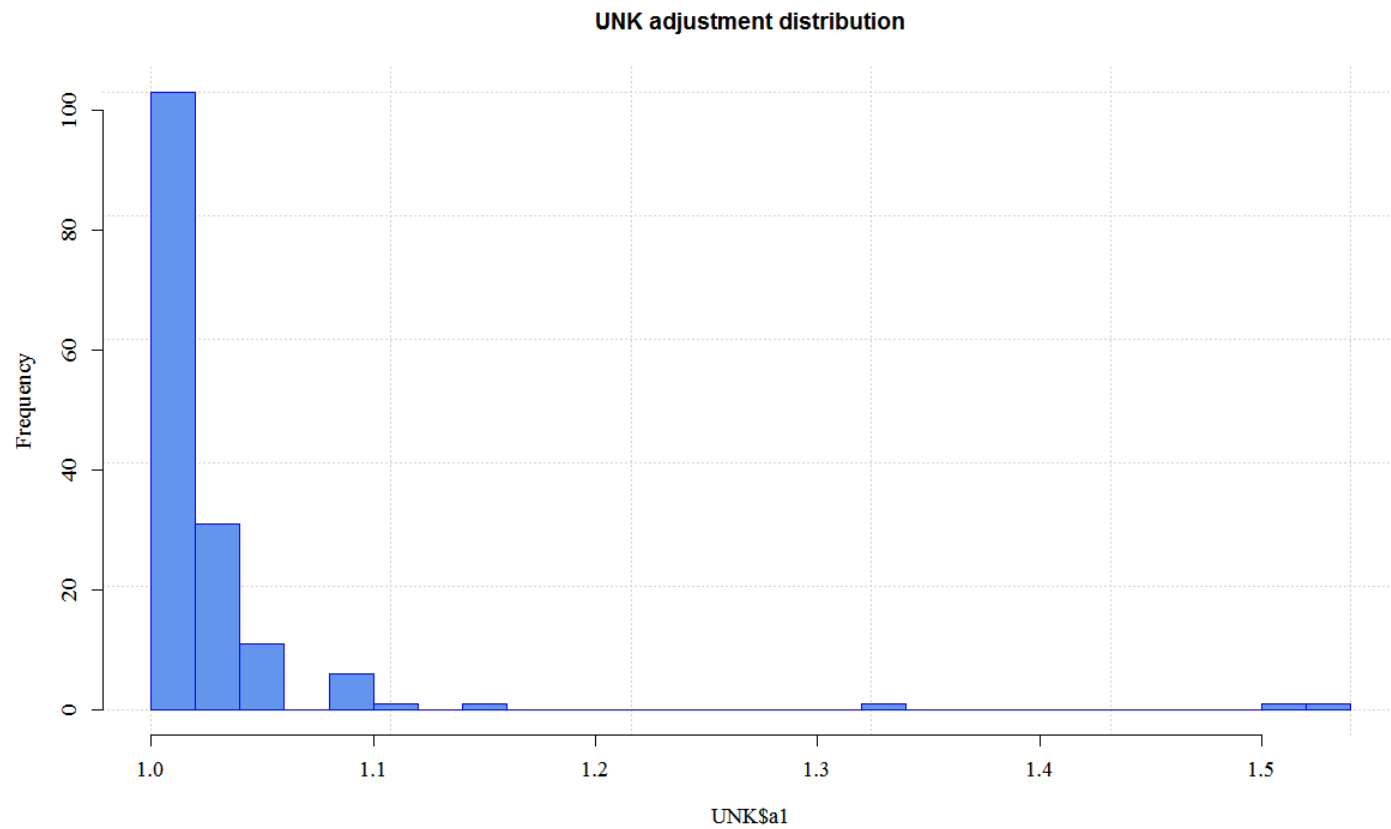
$$de_b = \frac{\sum_{i \in s_b} d_{0i}}{\sum_{i \in s_b, EC} d_{0i}}$$

3. El ponderador ajustado para la unidad $i \in s_{EC}$ es

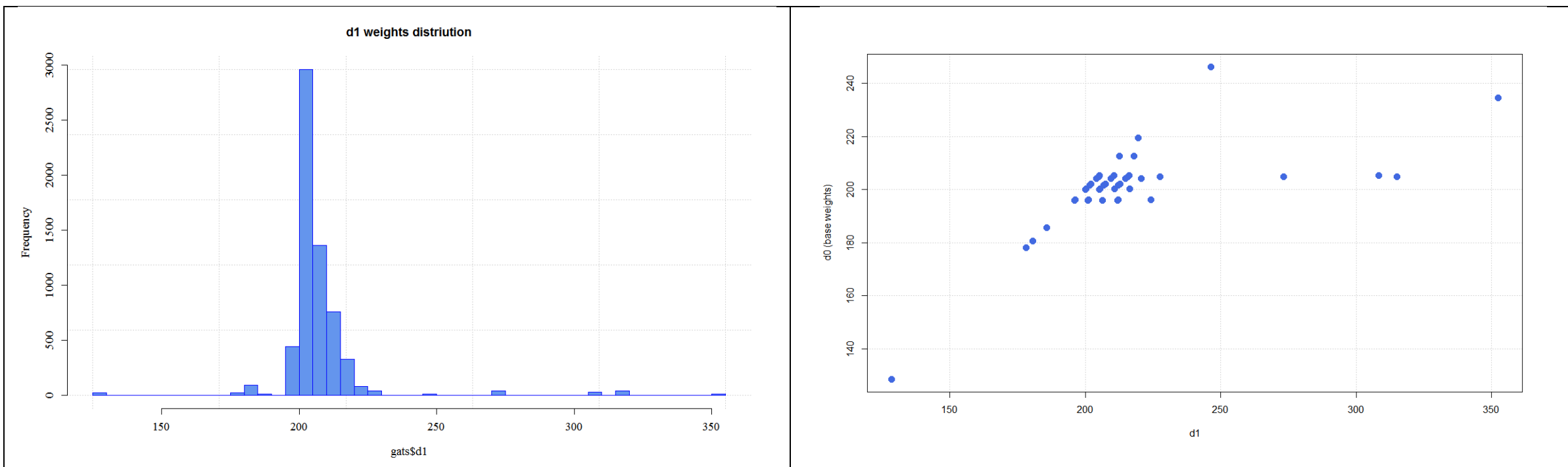
$$d_{1i} = de_b \times d_{0i}$$

Ajuste desconocimiento de elegibilidad GATS

Las clases=UPM



Ajuste desconocimiento de elegibilidad GATS



No respuesta

Bajo el supuesto de que todos los elementos tienen probabilidad ϕ_i de responder, el sesgo del estimador $\hat{y} = \sum_R d_{0i} \times y_i / \sum_R d_{0i}$ viene dado por

$$B(\hat{y}) = \frac{1}{N\bar{\phi}} \sum (y_i - \bar{Y})(\phi_i - \bar{\phi})$$

El sesgo depende de la covarianza entre la variable de interés y la probabilidad de responder.

Tipos de no respuesta

MCAR=la probabilidad de responder ϕ_i no dependen de las y 's ni las x 's. Todos tienen igual ϕ_i .

MAR= la probabilidad de responder ϕ_i no dependen de las y 's, pero si de las variables auxiliares x 's. Se puede modelar o estimar la probabilidad de responder utilizando las x 's, $\hat{\phi}(x_i)$.

NINR= la probabilidad de responder ϕ_i , depende de alguna de las variables de interés (y 's). Imposible modelar. No tenemos y para s_{ENR}

No respuesta

Si conociéramos las ϕ_i , podríamos un nuevo sistema de ponderadores $d_{2i} = \phi_i \times d_{0i}$ que produzcan estimaciones insesgadas.

Podemos intentar estimarlas, $\hat{\phi}_i$, dos formas:

1. Creando clases de ajuste de NR
2. Utilizando propensity score (logit, probit)
 - Propensity stratification
 - Árboles

Clases de ajuste por No respuesta

Idea: crear clases en donde, todas las unidades tengan la misma probabilidad de responder (e.i los mismos valores de las y 's)

- En la práctica obviamente las y 's son desconocidas.
- Construimos las clases en base a las TR, ponderadas o no.

El ponderador ajustado por NR en la clase c es

$$d_{2i} = NR_i \times d_{1i} = \frac{\sum_{s_{c,E}} d_{1i}}{\sum_{s_{c,ER}} d_{1i}} \times d_{1i}$$

Si tenemos muchas variables, debemos hacer las interacciones.

Propensity score

Podemos estimar $\hat{\phi}_i$ por medio de un probit o logit.

- Minimiza el sesgo si la NR es MAR o MCAR. **Lo es?**
- Esto nos permite utilizar varias variables, ya sea cualitativa o cuantitativa.
- Elegir si el modelo es ponderado o no.

Propensity score – ajuste unitario

Dos opciones de ajuste:

1. **Ajuste unitario.** Utilizamos un ajuste para cada $i \in s_{ER}$

$$d_{2i} = \frac{1}{\hat{\phi}(x_i)} \times d_{1i}$$

Observación: pueden agregarle variabilidad extra a los estimadores producto del aumento de la variabilidad de los ponderadores (calcular deff de kish)

Propensity score stratification

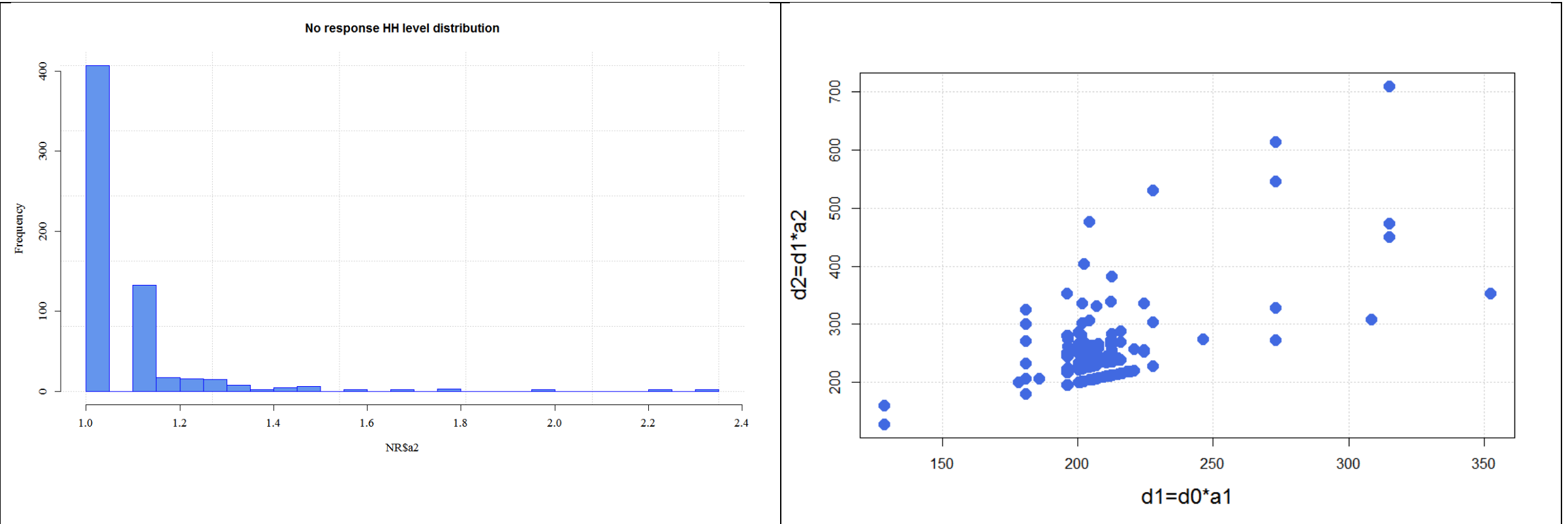
Propensity stratification: utilizamos los $\hat{\phi}(x_i)$ para crear clases y utilizamos un ajuste común para todos los respondentes dentro de la misma clase

1. Calculamos $\hat{\phi}(x_i)$ por medio del modelo
2. Ordenamos $\hat{\phi}(x_i)$ de menor a mayor.
3. Formamos clases con la misma cantidad de casos (R+NR)
 - i. Generalmente quintiles. Si n es “grande” más es OK.
 - ii. Tener cuidado la variabilidad en los $\hat{\phi}(x_i)$.

```
pclass(formula, data, link="logit",  
       numcl=5, type, design=NULL)
```

NR en GATS a nivel de hogar

Clases de Ajuste NR=UPM



NR en GATS

Strata	Description	deffK
1	Montevideo low	1.2092
	Montevideo medium	
2	low	1.0156
3	Montevideo médium	1.0063
4	Montevideo medium high	1.0070
5	Montevideo high	1.0276
6	Metropolitan ring	1.0235
7	Localities with more than 20,000 people	1.0060
8	Localities with 5,000 to 20,000 people	1.0190
9	Localities with 1,000 to 5,000 people	1.0288
10	Localities with less than 1,000 people	1.0550
11	Rural	1.0224

Ponderador a nivel de persona

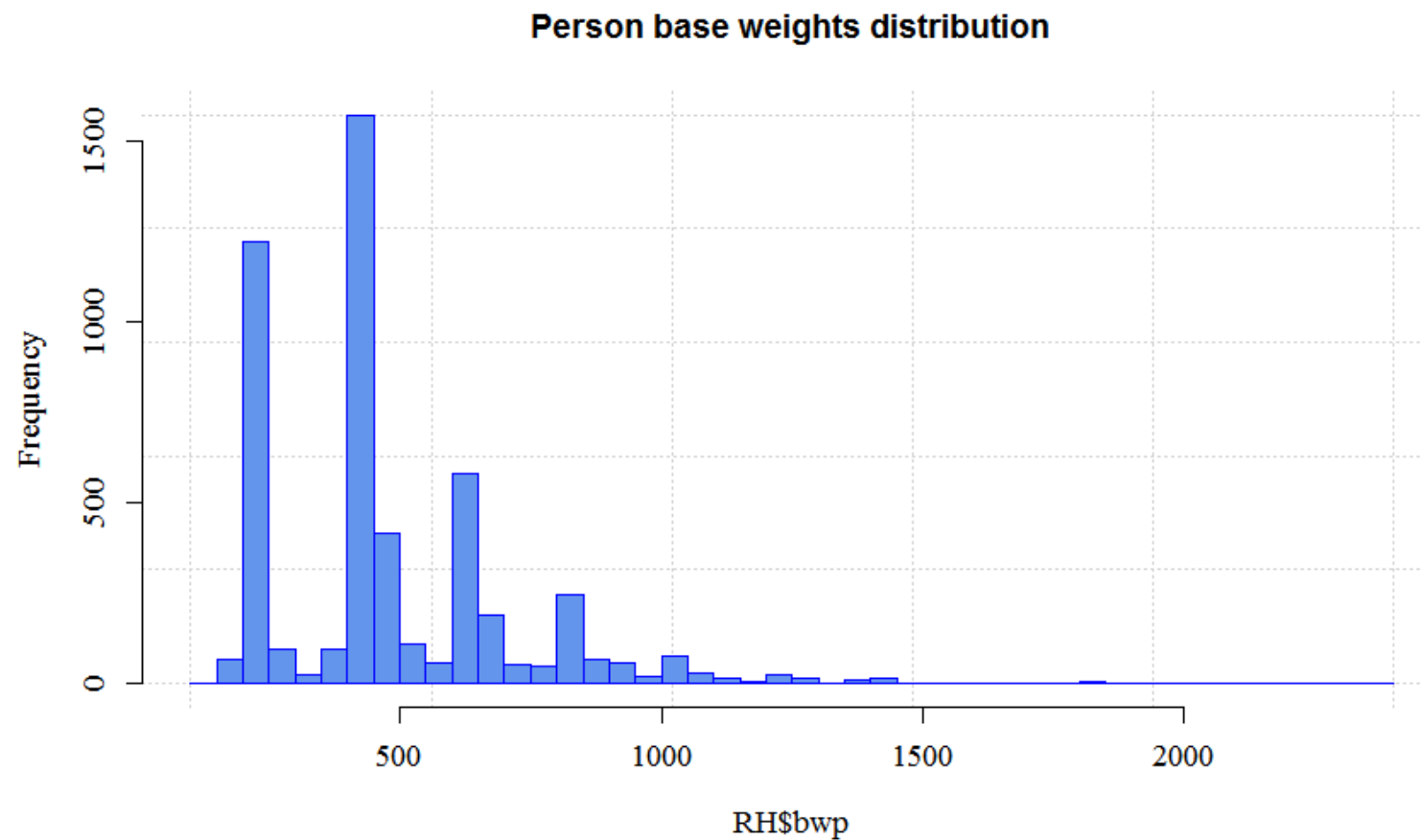
EL ponderador de la persona j perteneciente a la vivienda i

$$w_{ij} = d_{2i} \times P_i$$

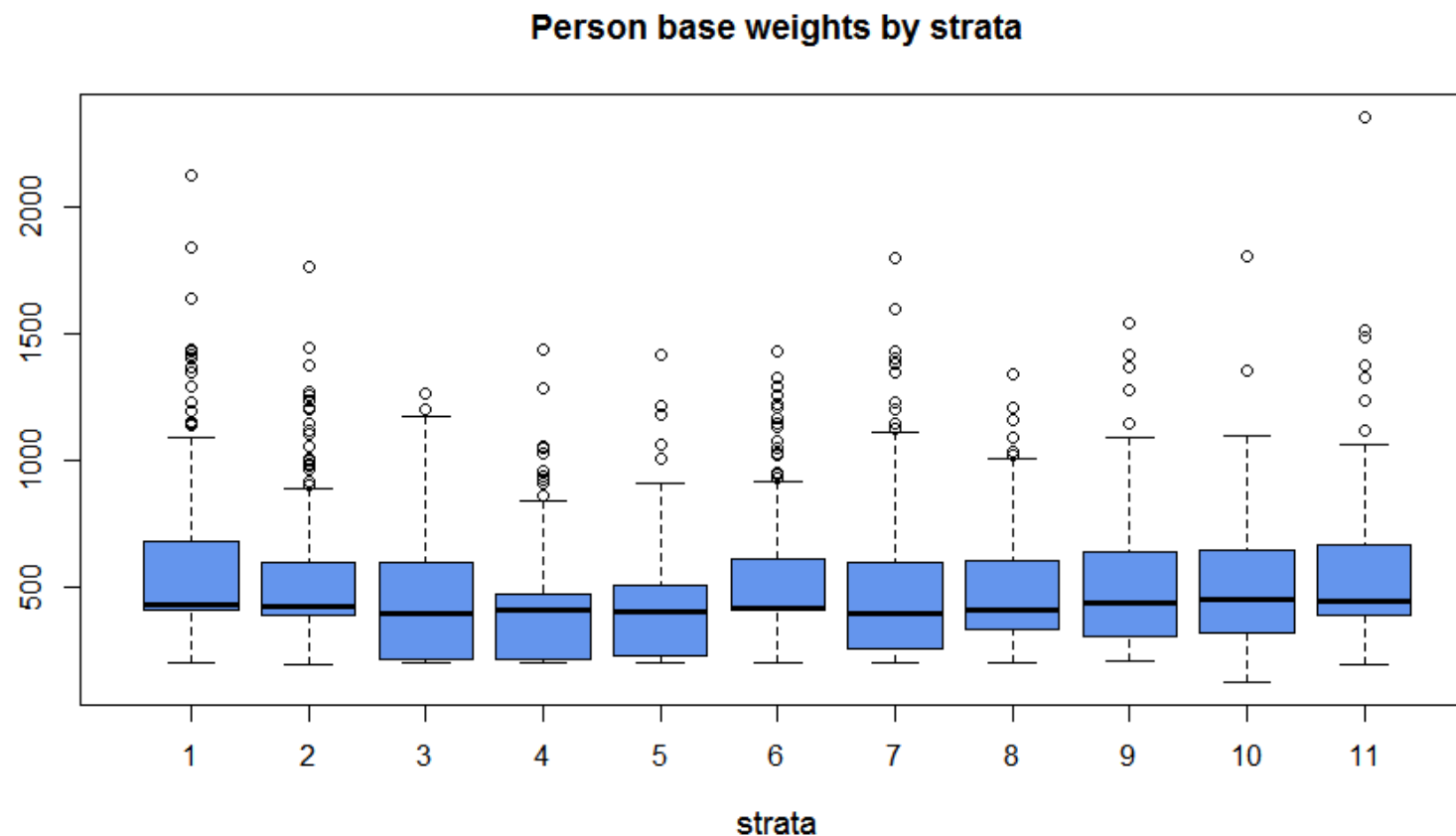
donde P_i es la cantidad de personas elegibles en la vivienda.

Strata	Description	Min.	1st Qu.	Median	Mean	3er Qu.	Max
1	Montevideo low	204.9	409.7	431.3	573.6	682.9	2128
2	Montevideo medium low	195.9	391.8	423.6	482.2	595.3	1763
3	Montevideo medium	200.2	216.5	400.5	430.4	600.7	1265
4	Montevideo medium high	205.3	216.1	410.5	410.1	473.7	1437
5	Montevideo high	202.2	227.4	404.4	430.4	500.7	1415
6	Metropolitan ring	204.2	408.3	418.8	486.8	612.5	1429
7	Localities with more than 20,000 people	200.1	256.8	400.3	455.6	600.4	1801
8	Localities with 5,000 to 20,000 people	201.6	336	413.6	476.2	604.8	1344
9	Localities with 1,000 to 5,000 people	212.6	308.4	436.1	517.5	637.8	1542
10	Localities with less than 1,000 people	128.5	321.2	451.8	516.1	650.5	1804
11	Rural	196.2	392.4	447.2	533.2	670.7	2354

Ponderador a nivel de persona



Ponderador a nivel de persona



Ponderador a nivel de persona

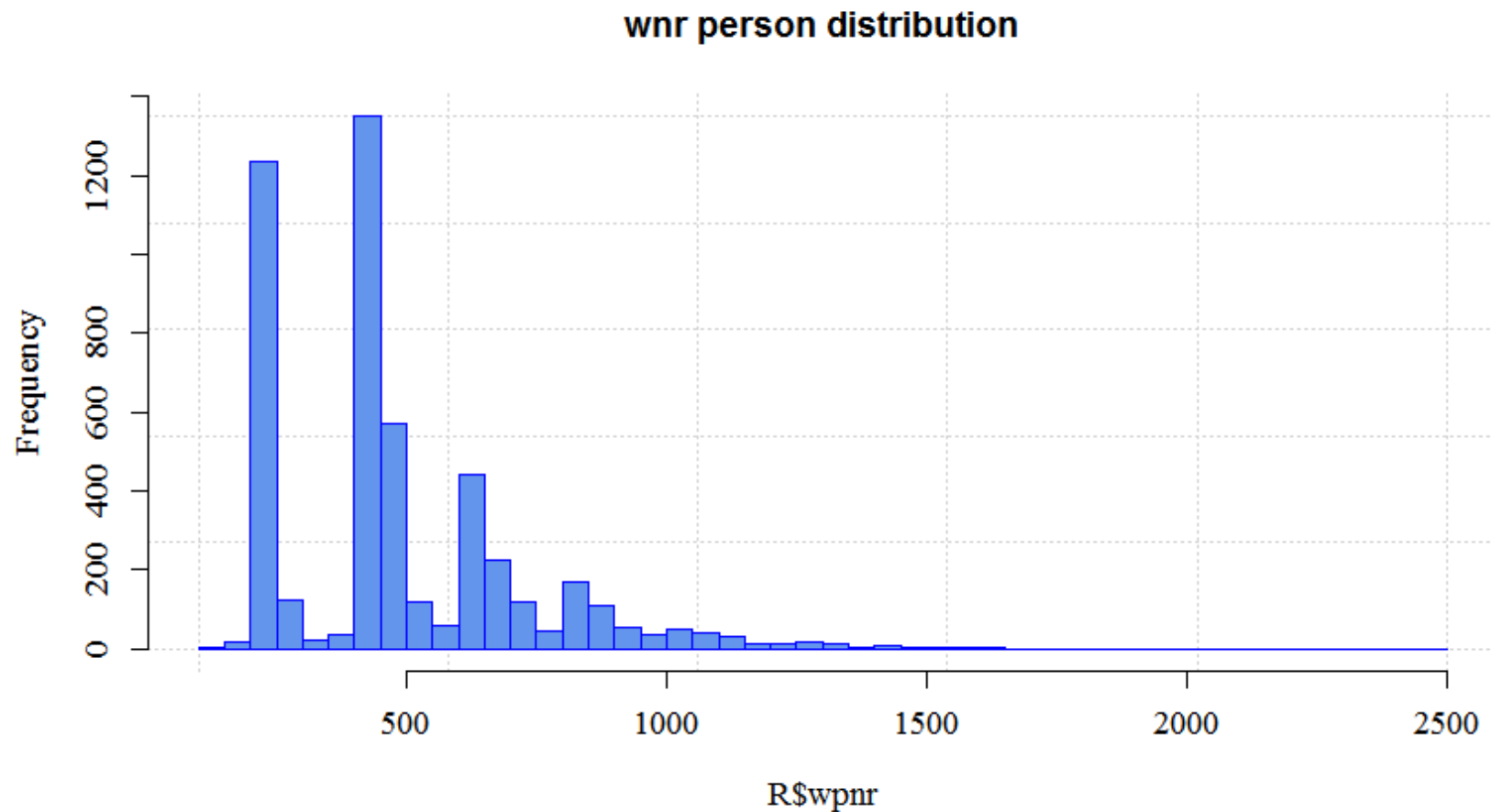
Strata	Description	deffK
1	Montevideo low	1.398914
2	Montevideo medium low	1.270543
3	Montevideo medium	1.243899
4	Montevideo medium high	1.235167
5	Montevideo high	1.253567
6	Metropolitan ring	1.237132
7	Localities with more than 20,000 people	1.225894
8	Localities with 5,000 to 20,000 people	1.214831
9	Localities with 1,000 to 5,000 people	1.244486
10	Localities with less than 1,000 people	1.277887
11	Rural	1.300022

Ajuste no respuesta a nivel de persona

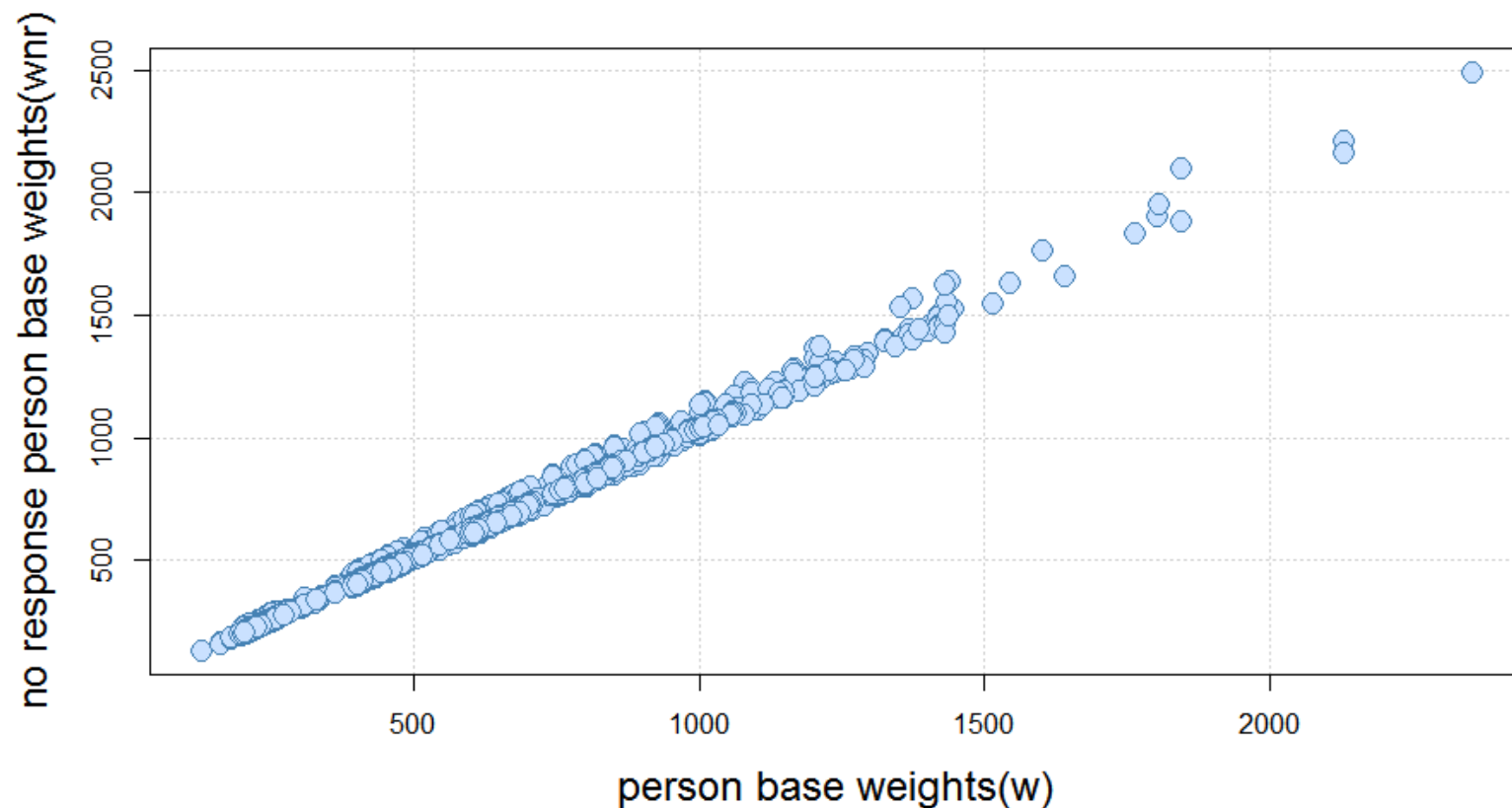
Clases de Ajuste NR=SEXO × EDAD × FUMA

		Male		Female	
	Age	a3=1/RR	weighted response rate (RR)	a3=1/RR	weighted response rate (RR)
Non smokers	[15,25)	1.0585	0.9447	1.0419	0.9597
	[25,35)	1.0251	0.9755	1.0528	0.9498
	[35,45)	1.0304	0.9705	1.0147	0.9855
	[45,55)	1.0248	0.9758	1.0234	0.9772
	[55,Inf)	1.0462	0.9558	1.0402	0.9614
Smokers	[15,25)	1.1407	0.8767	1.1005	0.9087
	[25,35)	1.1004	0.9087	1.0298	0.9710
	[35,45)	1.0841	0.9225	1.0721	0.9328
	[45,55)	1.1352	0.8809	1.0000	1.0000
	[55,Inf)	1.0137	0.9865	1.0176	0.9827

Ajuste no respuesta a nivel de persona



Ajuste no respuesta a nivel de persona



Ajuste no respuesta a nivel de persona

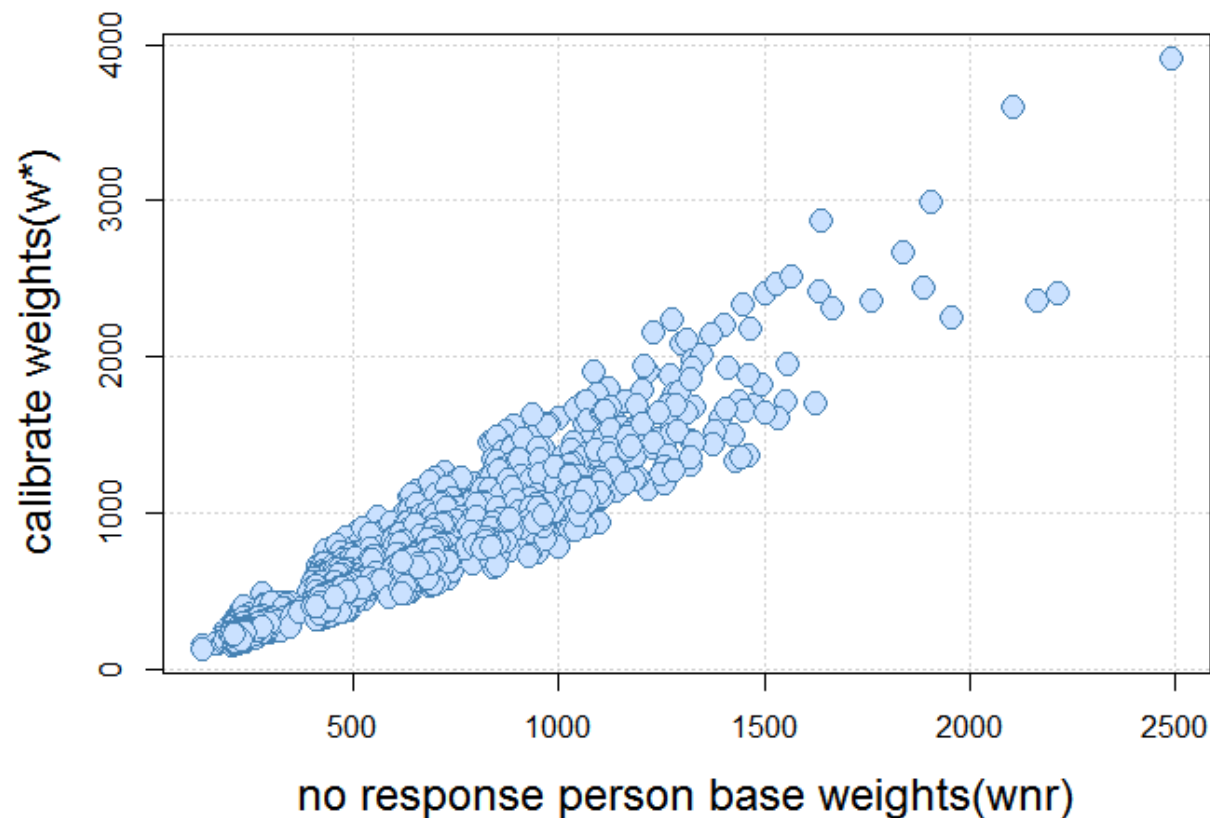
Strata	Description	deffK
1	Montevideo low	1.393
2	Montevideo medium low	1.27508
3	Montevideo medium	1.25154
4	Montevideo medium high	1.23728
5	Montevideo high	1.26092
6	Metropolitan ring	1.24316
7	Localities with more than 20,000 people	1.23468
8	Localities with 5,000 to 20,000 people	1.22347
9	Localities with 1,000 to 5,000 people	1.25197
10	Localities with less than 1,000 people	1.29246
11	Rural	1.30807

Calibración a conteos poblacionales

- Raking
- Variables de control:
 - Sexo con interacción del tramo de edad (15 a 24, 25 a 34, 35 a 44, 45 a 54, 55 o +)
 - Nivel educativo (Primaria, Primer Ciclo, Bachillerato y terciario)
 - Region (Montevideo e Interior)

Los conteos poblacionales provienen de la ECH

Calibración a conteos poblacionales



Calibración a conteos poblacionales

