

Els mètodes de captura-recaptura: creuant llistes

Pere Puig

Pere.Puig@uab.cat

Department de Matemàtiques

Universitat Autònoma de Barcelona

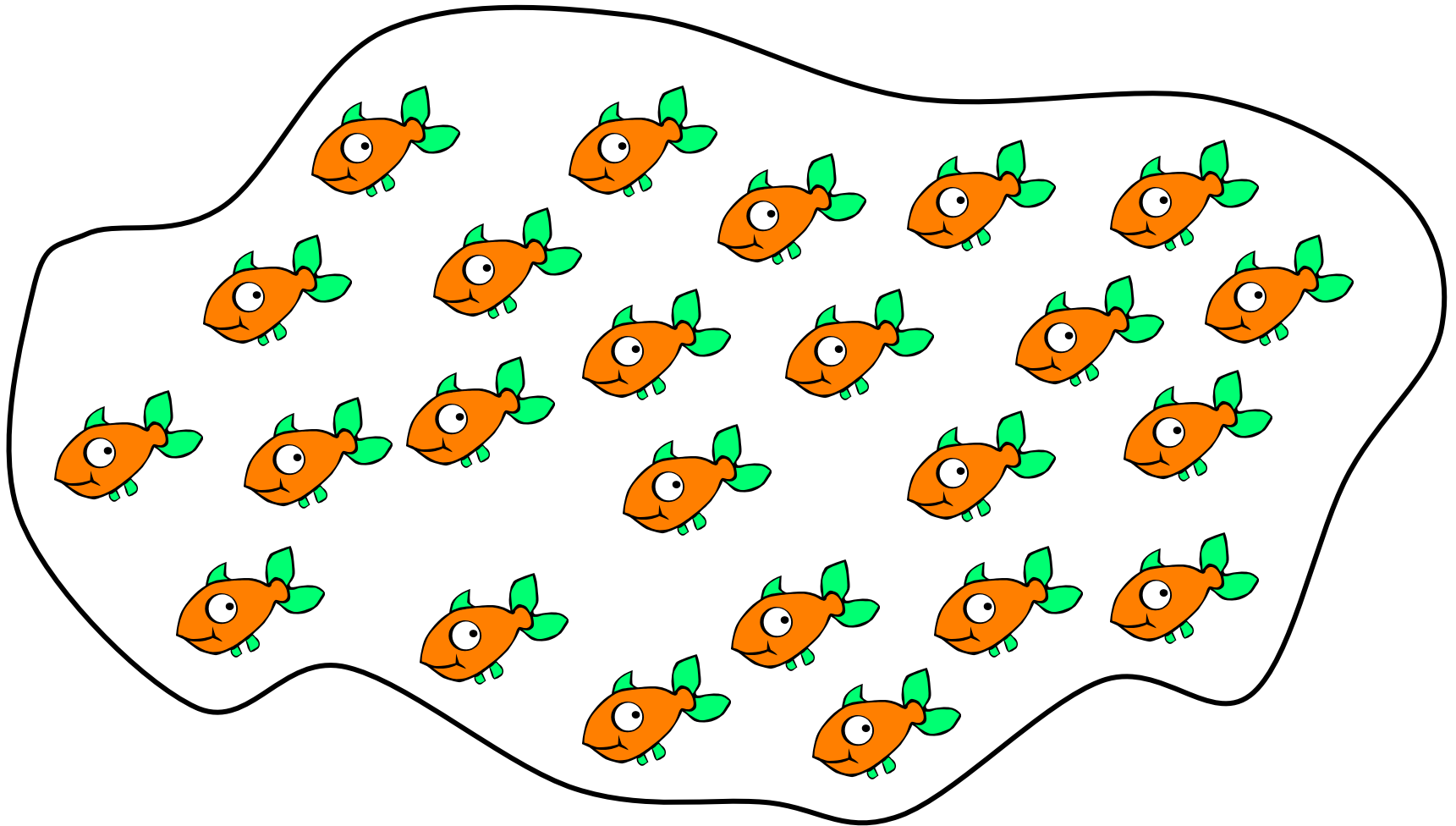
Advanced Stochastic Modelling Research Group

Poblacions amagades o “invisibles”

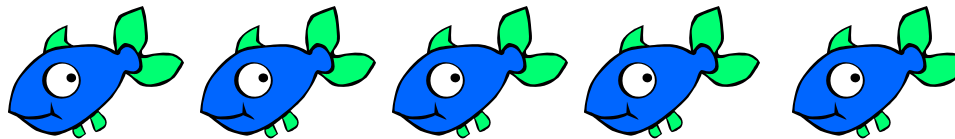
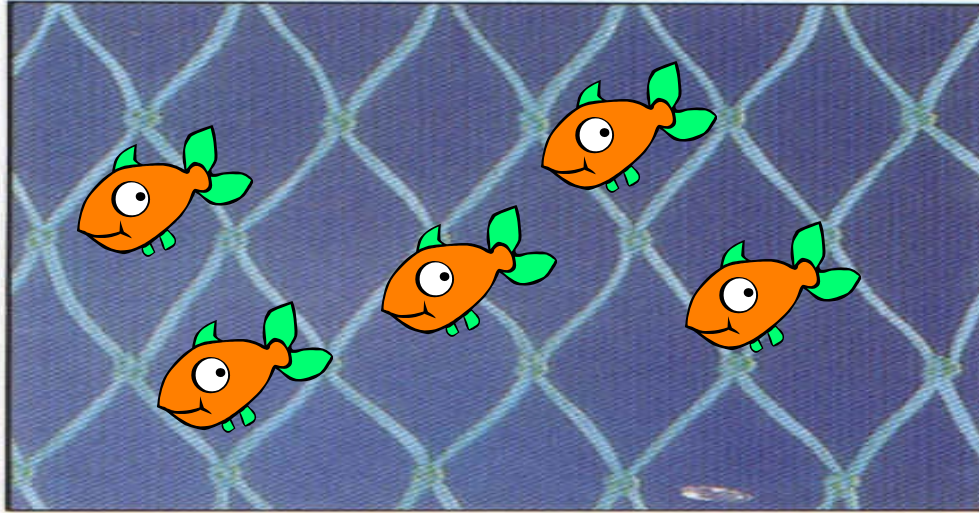
- No existeix un mètode directe de mostreig.
- L'accés als membres de la població és complicat.
- El reconeixement públic de la pertinença a la població és potencialment difícil.

Els mètodes de captura-recaptura

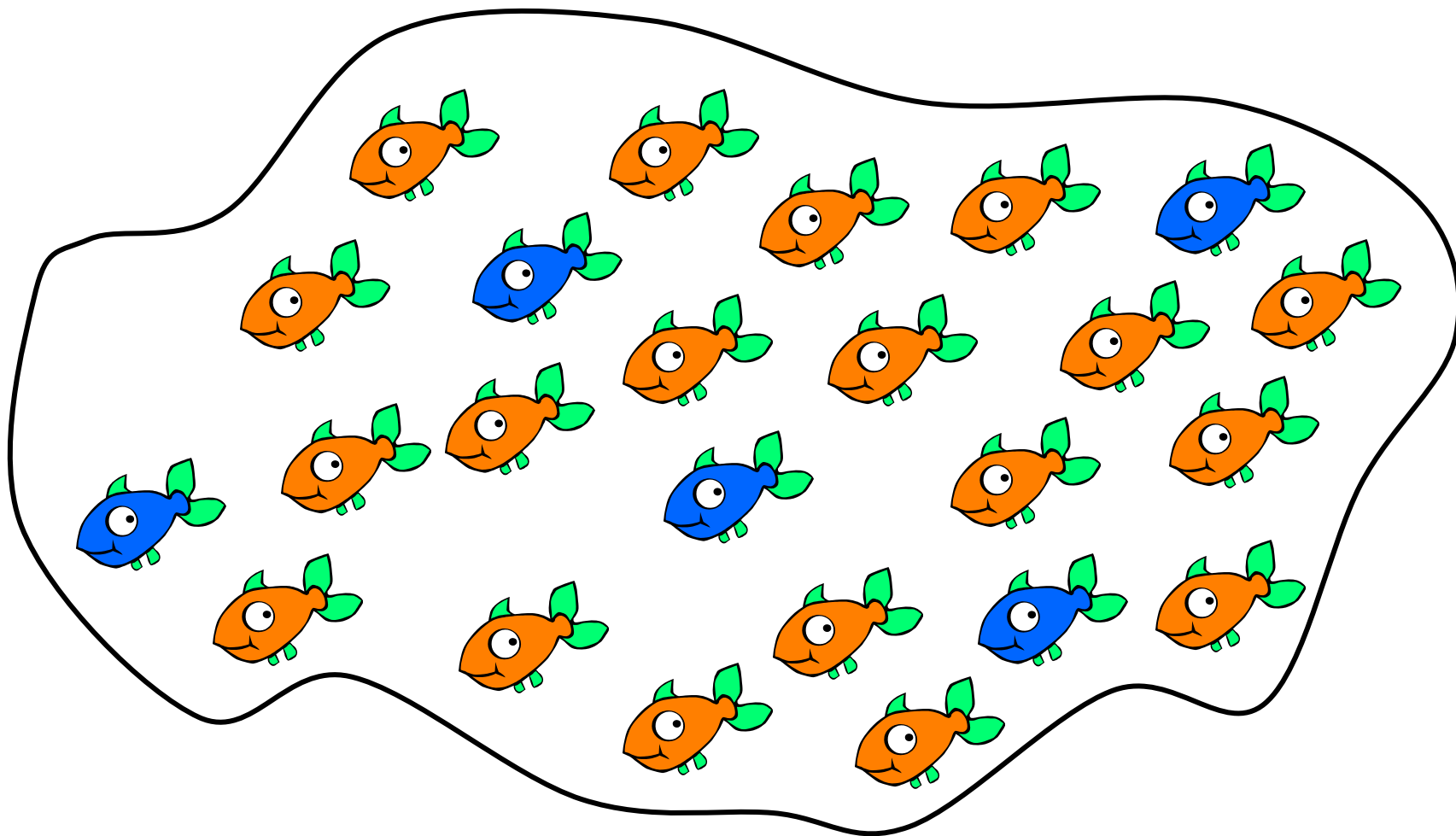
Quants peixos hi ha en aquest estany ?



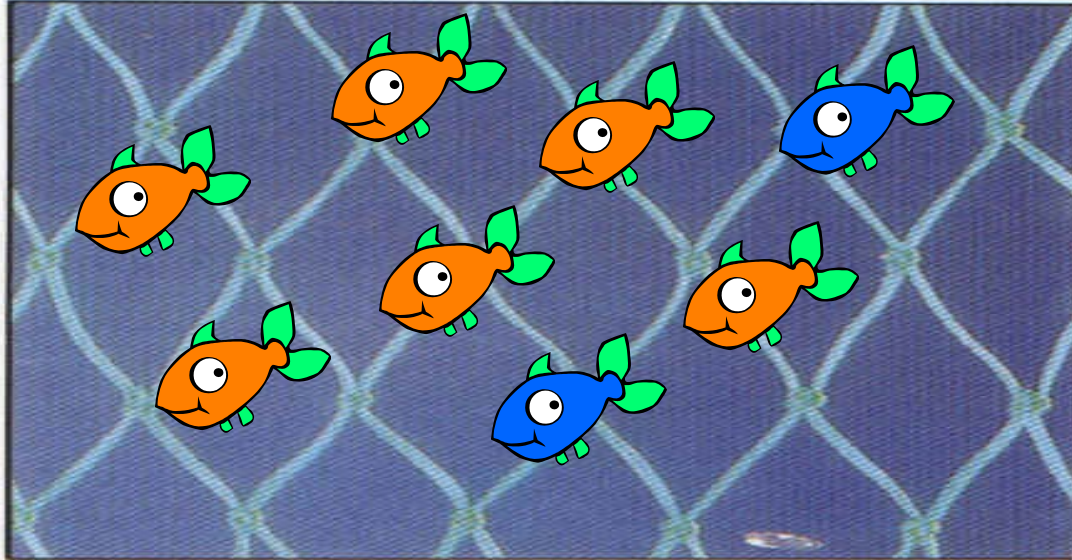
Capturem uns quants i els marquem



Els marcats queden repartits homogèniament
entre els altres.



Tornem a fer una segona captura



Hem capturat 8, dels quals 2 estan marcats.

N : nombre total de peixos en el llac (?)

n₁ : nombre de peixos de la primera captura
(i posteriorment marcats) (5)

La proporció de peixos que hi ha marcats en total

és: $\frac{n_1}{N}$

n₂ : nombre de peixos de la segona captura (8)

m: nombre de peixos que trobem marcats en
aquesta segona captura (2)

La proporció de peixos que trobem marcats en

aquesta segona captura és: $\frac{m}{n_2}$

$$\frac{n_1}{N} \approx \frac{m}{n_2}$$

$$N \approx \frac{n_1 n_2}{m}$$

$$\hat{N} = \frac{n_1 n_2}{m}$$

Això és el que es coneix actualment com l'estimador de Lincoln-Petersen de la grandària poblacional.

Pel nostre exemple,

$$n_1 = 5$$

$$n_2 = 8$$

$$m = 2$$

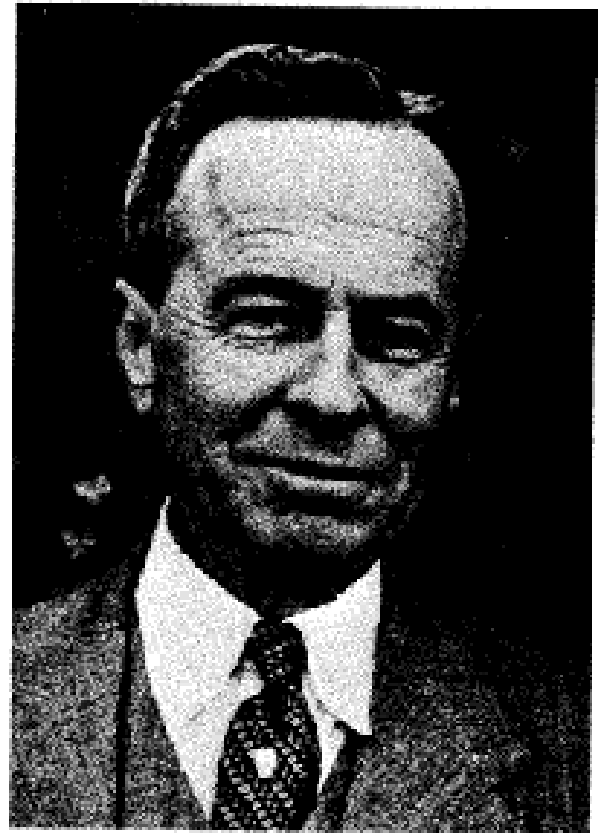
$$\hat{N} = \frac{5 \times 8}{2} = 20$$

<https://www.youtube.com/watch?v=Hi460--lUhm&t=3s>

Els divulgadors del mètode



Carl George Johannes Petersen



Frederick C. Lincoln

Condicions bàsiques del model

- La població és tancada.
- Cada mostra és aleatòria.
- Tots els animals tenen la mateixa probabilitat de ser capturats a cada mostra.
- Captura i marcatge no afecten a la probabilitat de recaptura.
- Les marques no desapareixen o es perden.

L'inventor



Laplace, el 1783, va utilitzar aquest mètode per estimar el nombre d'habitants de França.

Un cens incomplet va ser la primera captura (n_1).

Els individus d'una enquesta van constituir la segona (n_2).

El nombre individus coincidents en ambdues llistes o fonts va ser m .

Chapman & Hall/CRC
Interdisciplinary Statistics Series

Capture-Recapture Methods for the Social and Medical Sciences

Edited by
Dankmar Böhning
Peter G.M. van der Heijden
John Bunge

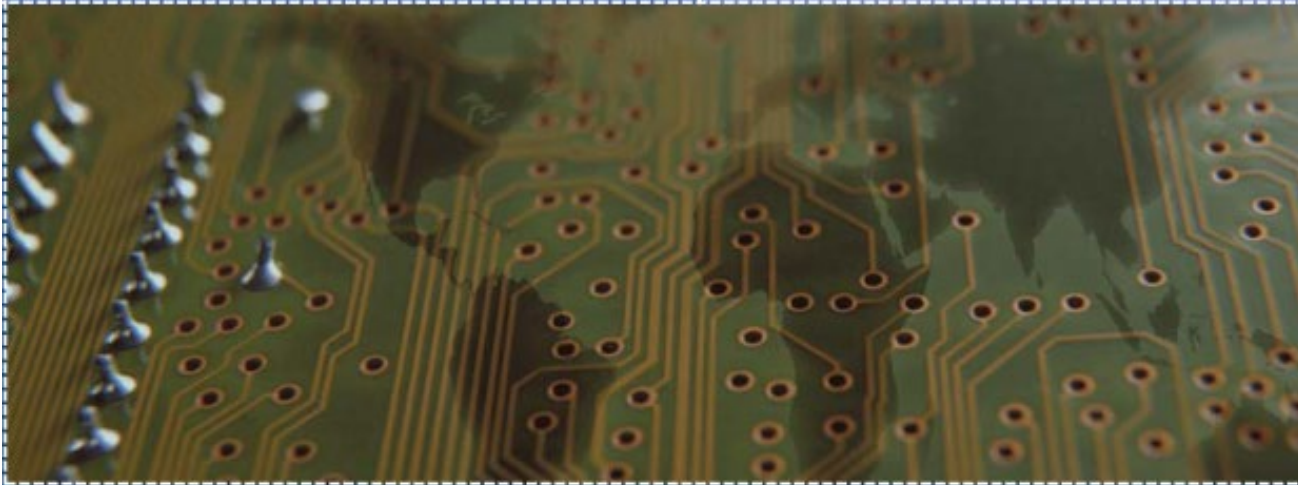
 CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK



VISIÓN ESTRATÉGICA DE UNODC PARA AMÉRICA LATINA Y EL CARIBE



#UNODC2025



Estimación de la prevalencia —
Métodos indirectos para
estimar la magnitud
del problema de las drogas

<i>Fuente de datos</i>	<i>Ejemplo</i>
Tratamiento especializado contra las drogas	Drogadictos en régimen de metadona, que acuden a centros de tratamiento o internados en un centro asistencial
Organismos de terapia contra la toxicomanía, bajo umbral	Drogadictos que acuden a ambulatorios o son contactados por personal de acción exterior
Programas de intercambio de agujas	Drogadictos registrados en programas de intercambio de agujas
Salas de urgencias	Drogadictos internados en salas de urgencias por sobredosis
Laboratorios	Drogadictos que se someten a pruebas de detección del VIH, del VHC o del virus de la hepatitis B
Policía/prisiones	Drogadictos arrestados o encarcelados por delitos de drogas o por otros delitos
Servicios de libertad condicional	Drogadictos en libertad condicional
Evaluaciones efectuadas por servicios sociales	Drogadictos evaluados por servicios sociales locales
Albergues para drogadictos	Drogadictos que viven en albergues
Registros de toxicómanos	Drogadictos notificados a un registro central
Encuestas sobre drogadictos problemáticos	Encuestas comunitarias sobre drogadicción
Muertes por sobredosis	Número de muertes por sobredosis de opiáceos

L'exemple de Bangkok

(AJPH, 1994)

Estimating the Number of HIV-Infected Injection Drug Users in Bangkok: A Capture–Recapture Method

Timothy D. Mastro, MD, Dwip Kitayaporn, MD, DrPH, Bruce G. Weniger, MD, MPH, Suphak Vanichseni, MD, MPH, PhD, Vuti Laosunthorn, PhD, Thongchai Uneklabh, MD, Chintra Uneklabh, MD, Kachit Choopanya, MD, MPH, and Khanchit Limpakarnjanarat, MD, MPH

Llista de consumidors d'opiacis inscrits en programes de tractament amb metadona. Dades d'assistència registrades a l'abril i maig de 1991: **$n_1 = 4064$**

Llista de persones que van donar resultat positiu en anàlisi d'orina d'opiacis, arrestades a les comissaries de Bangkok entre juny i setembre de 1991: **$n_2 = 1540$**

Persones en ambdues llistes: **$m = 171$**

$$\hat{N} = \frac{4064 \times 1540}{171} \sim 36600$$

TABLE 3—Estimates of the Number of Opiate Users in Bangkok (n), by Unstratified and Age-Stratified Analyses

Age Group	Capture 1	Capture 2	No. of Matches	n	95% Confidence Interval
Unstratified analysis					
All	4064	1540	171	36 600	31 538, 41 662
Stratified analysis					
< 20 y	76	78	4	1482	105, 2859
20–29 y	1378	699	76	12 674	10 059, 15 289
30–39 y	2106	644	80	16 953	13 543, 20 363
> 39 y	491	118	11	5267	2336, 8198
All	4051	1539	171	36 376	30 996, 41 757

THE INTERNATIONAL
JOURNAL
Of **SCIENCE**
IN SOCIETY

Using Capture-recapture to Estimate the Prevalence of Intimate Partner Violence: The Gender Symmetry Debate

Lori A. Post, Yale University, CT, USA

Nancy J. Mezey, Monmouth University, New Jersey, USA

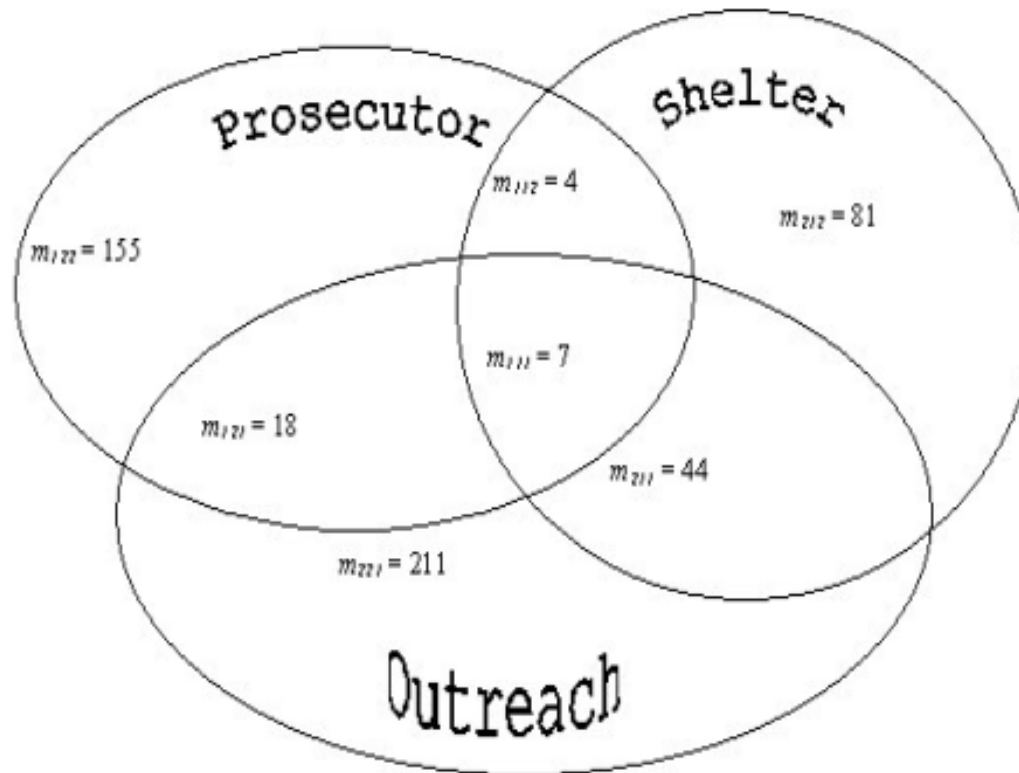
Christopher D. Maxwell, Michigan State University, MI, USA

Karin Rhodes, University of Pennsylvania School of Medicine, PA, USA

Using Capture-recapture to Estimate the
Prevalence of Intimate Partner Violence: The
Gender Symmetry Debate

Els autors consideren les coincidències en tres llistes:

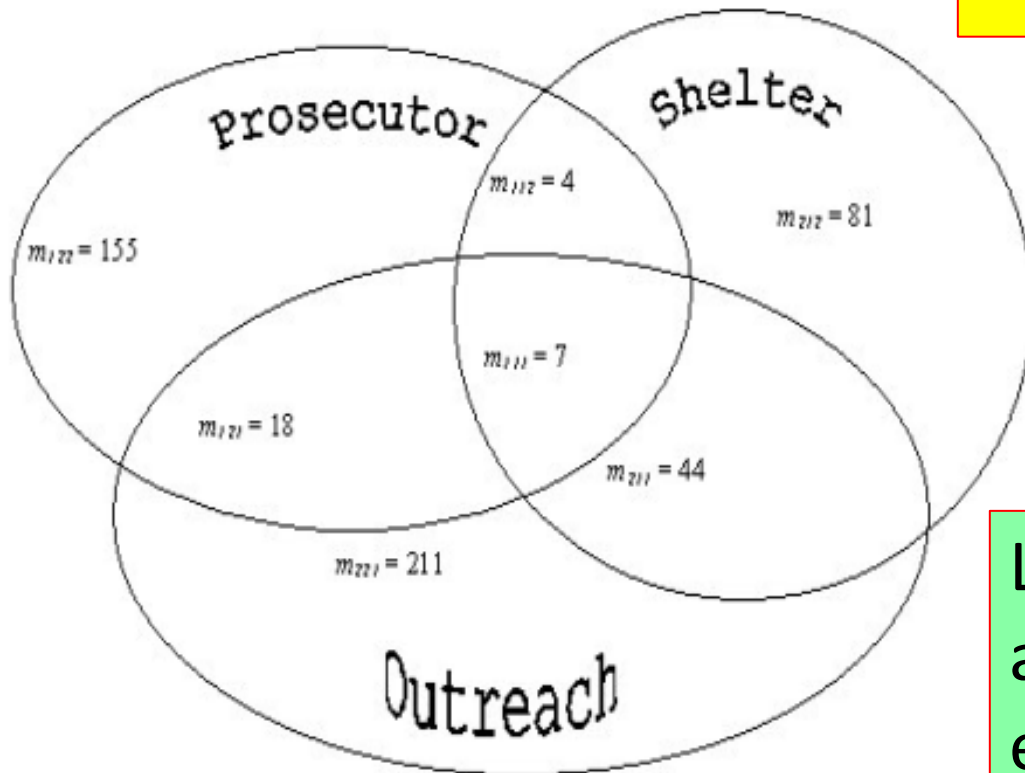
- **Prosecutor's** office (fiscalia)
- Local domestic violence **shelter** (refugis)
- Family **Outreach** agencies (serveis d'assistència familiar)



Considerem només 2 llistes:

- **Prosecutor** (fiscalia) = 155+4+18+7 = 184
- **Outreach** (assistència familiar) = 211+44+7+18 = 280
- En ambdues llistes = 18+7 = 25

$$\hat{N} = \frac{184 \times 280}{25} \sim 2061$$



L'estimació que fan els autors utilitzant les 3 llistes és de 2337 dones.

Problemes al treballar amb llistes

Heterogeneïtat: Diferents individus tenen diferents probabilitats de ser “capturats”. Els valors poden dependre del sexe, edat, situació social, etc.

Dependència: El fet de que un individu estigui en una llista afecta a la probabilitat de que estigui en una altra.

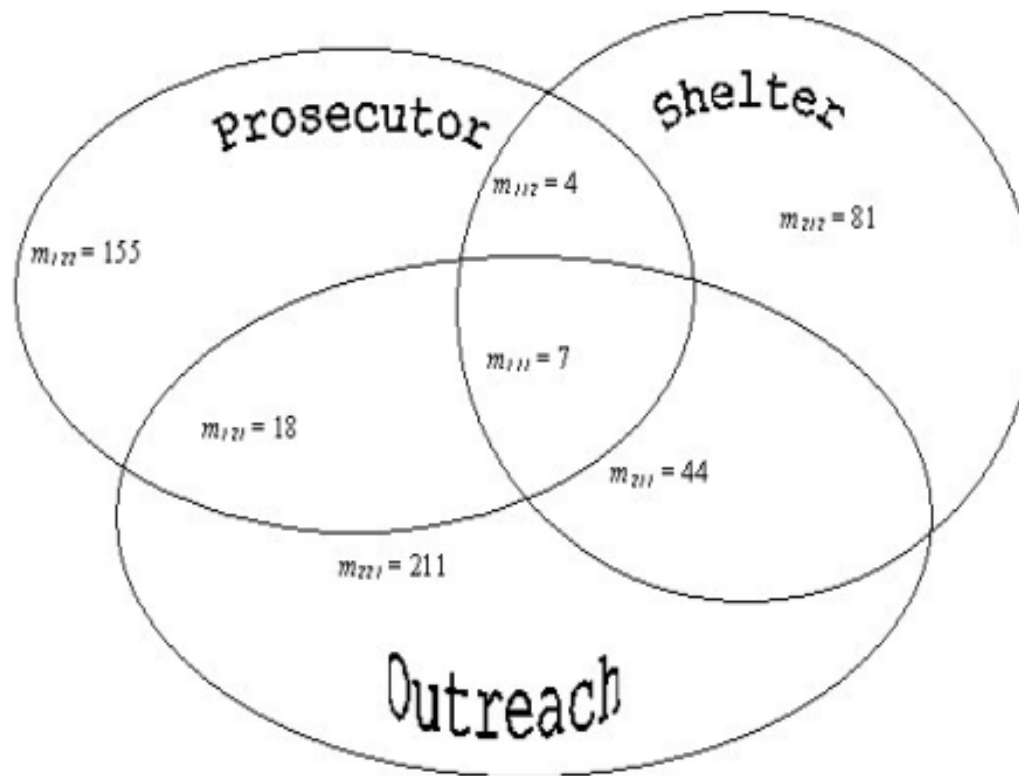
Possibles solucions

Heterogeneïtat: Estratificació.

Considerar un model independent per cada agrupació o estrat.

Dependència: Models més complicats.

Per exemple els models log-lineals.



$$\log(E(m_{ijk})) = a + bL_1 + cL_2 + dL_3 + eL_1L_2 + fL_1L_3 + gL_2L_3$$

L_i són les variables indicadores de les llistes i a, b, c, \dots els paràmetres del model. Observem que $\log(E(m_{000})) = a$.

$$\log(E(N))=a+bPr+cSh+dOu+ePrSh+fPrOu+gShOu$$

N	Pr	Sh	Ou
155	1	0	0
4	1	1	0
18	1	0	1
7	1	1	1
81	0	1	0
44	0	1	1
211	0	0	1

```

glm(formula = N ~ Pr + Sh + Ou + Pr:Ou + Pr:Sh + Sh:Ou, family = poisson,
     data = data)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   8.6748     0.7033  12.334 < 2e-16 ***
Pr            -3.6314     0.6987  -5.197 2.02e-07 ***
Sh            -4.2803     0.6945  -6.163 7.13e-10 ***
Ou            -3.3229     0.7000  -4.747 2.06e-06 ***
Pr:Ou          1.1699     0.6542   1.788  0.0737 .
Pr:Sh          0.6232     0.4753   1.311  0.1898
Sh:Ou          2.7127     0.6744   4.022 5.77e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 5.2901e+02  on 6  degrees of freedom
Residual deviance: 1.0880e-14  on 0  degrees of freedom
AIC: 51.744

```

$$m_{000} \sim \exp(8.6748) = 5853.53$$

$$\log(E(N)) = a + bPr + cSh + dOu + fPrOu + gShOu$$

```
glm(formula = N ~ Pr + Sh + Ou + Pr:Ou + Sh:Ou, family = poisson,  
     data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	8.0516	0.5185	15.530	< 2e-16	***
Pr	-3.0082	0.5122	-5.873	4.28e-09	***
Sh	-3.6571	0.5064	-7.222	5.14e-13	***
Ou	-2.7114	0.5230	-5.184	2.17e-07	***
Pr:Ou	0.6858	0.5534	1.239	0.215	
Sh:Ou	2.1552	0.5296	4.070	4.70e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 529.006 on 6 degrees of freedom
Residual deviance: 1.592 on 1 degrees of freedom
AIC: 51.336

$$m_{000} \sim \exp(8.0516) = 3138.8$$

$$\log(E(N))=a+bPr+cSh+dOu+gShOu$$

```
glm(formula = N ~ Pr + Sh + Ou + Sh:Ou, family = poisson, data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	7.4932	0.2095	35.76	< 2e-16	***
Pr	-2.4498	0.1935	-12.66	< 2e-16	***
Sh	-3.1334	0.2235	-14.02	< 2e-16	***
Ou	-2.1423	0.2063	-10.38	< 2e-16	***
Sh:Ou	1.6315	0.2719	6.00	1.97e-09	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 529.0064 on 6 degrees of freedom
Residual deviance: 3.3617 on 2 degrees of freedom
AIC: 51.106

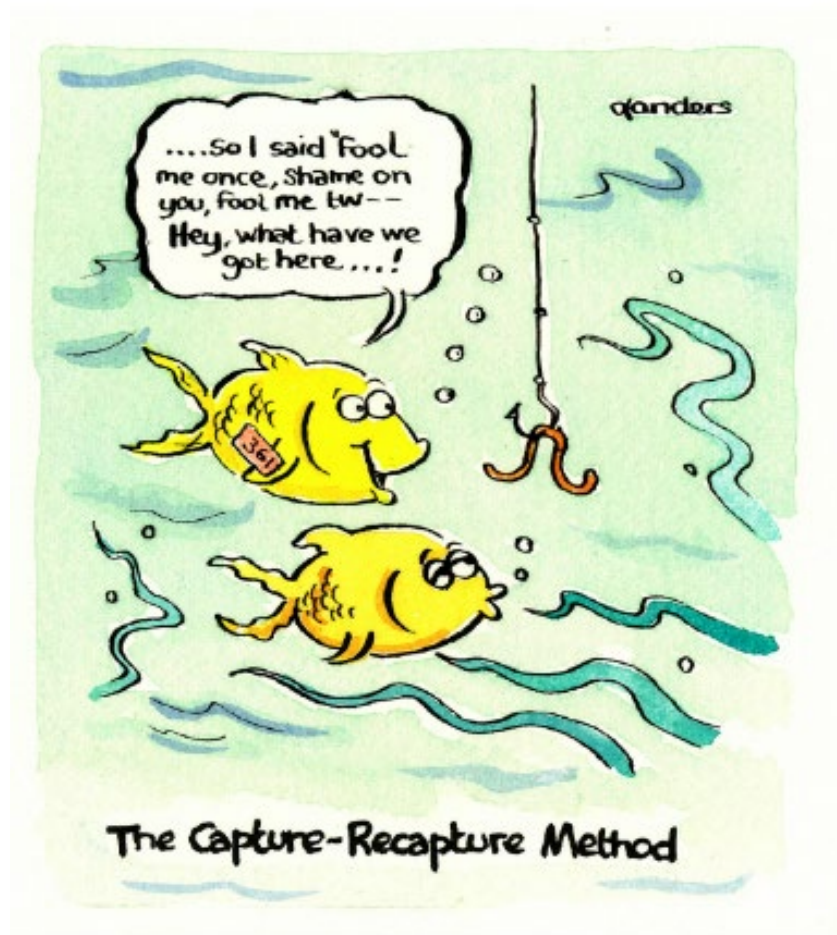
Number of Fisher Scoring iterations: 4

$$m_{000} \sim \exp(7.4932) = 1795.8$$

IC 95%: [1191.0, 2707.6]

$$N \sim 1796 + 520 = 2316$$

IC 95%: [1711,
3228]



Gràcies per la vostra atenció...