# aggregate association index

## & New Zealand election 1893

**Duy Tran[1],** A/Prof. Eric Beh[1], Prof. Irene Hudson[1], Linda Moore[2]

*1. University of Newcastle, Australia*

*2. Statistics New Zealand*

1 why **New Zealand** election

2 what is **Aggregate Association Index** (AAI)

3 what is the **role of AAI** in the NZ election

4 where am I **going** to from here

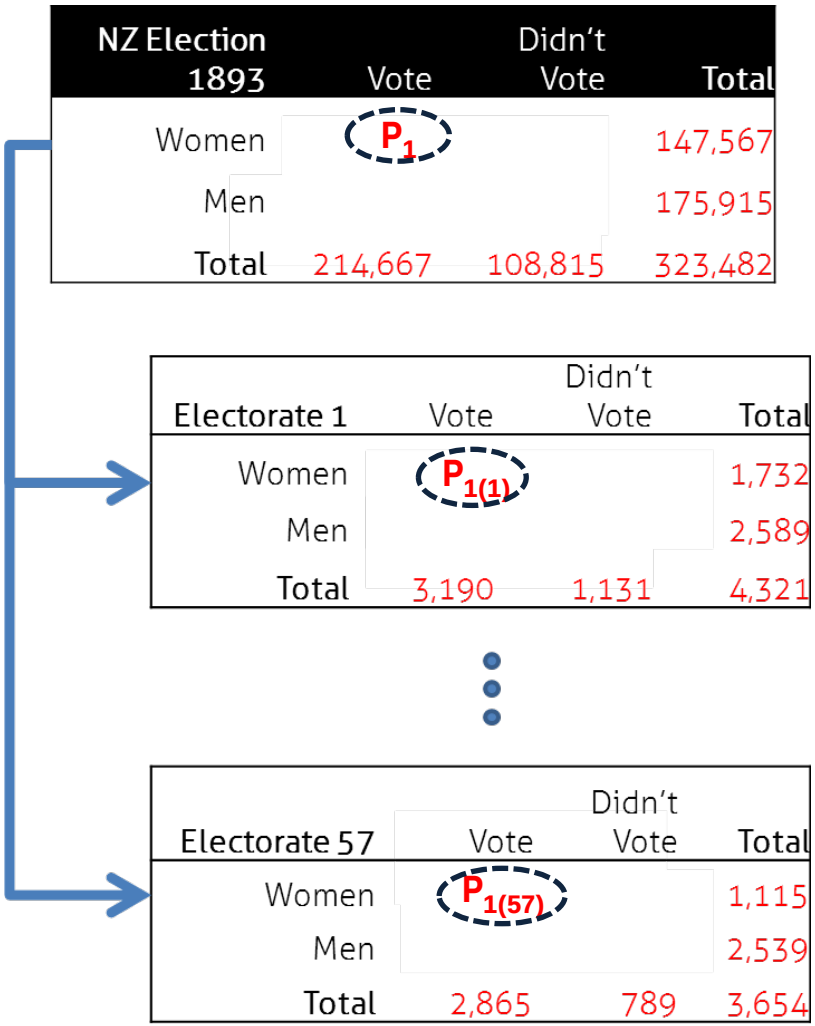

'Lady voters' approach the polling booth at the Drill Hall in Rutland Street, just off Queen Street Auckland, 6 Dec 1899.

| NZ Election 1893 | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | | | 147,567 |
| Men | | | 175,915 |
| Total | 214,667 | 108,815 | 323,482 |

| Electorate 1 | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | | | 1,732 |
| Men | | | 2,589 |
| Total | 3,190 | 1,131 | 4,321 |

| Electorate 57 | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | | | 1,115 |
| Men | | | 2,539 |
| Total | 2,865 | 789 | 3,654 |

# Ecological Fallacy




# Ecological Inference Techniques

*Hudson, Moore, Beh & Steel (2010, JRSSA)*

# Assumptions !

# Aggregate Association Index

{AAI

Proposed by BEH

(2008, JSPI; 2010,CSDA)

A single 2x2 table

with only marginal totals

Determine the possibility of an association

between 2 categorical variables

# Electorate 1, 1893

| Electorate 1 | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | | | 1,732 |
| Men | | | 2,589 |
| Total | 3,190 | 1,131 | 4,321 |

| Electorate g | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | $p_{11g}$ | $p_{12g}$ | $p_{1.g}$ |
| Men | $p_{21g}$ | $p_{22g}$ | $p_{2.g}$ |
| Total | $p_{.1g}$ | $p_{.2g}$ | 1 |

$$L_{P_{1g}} = max\left(0, \frac{n_{.1g} - n_{2.g}}{n_{1.g}}\right)$$

Define $\frac{n_{11g}}{n_{1.g}} = P_{1g}$

$$min\left(\frac{n_{.1g}}{n_{1.g}}, 1\right) = U_{P_{1g}}$$

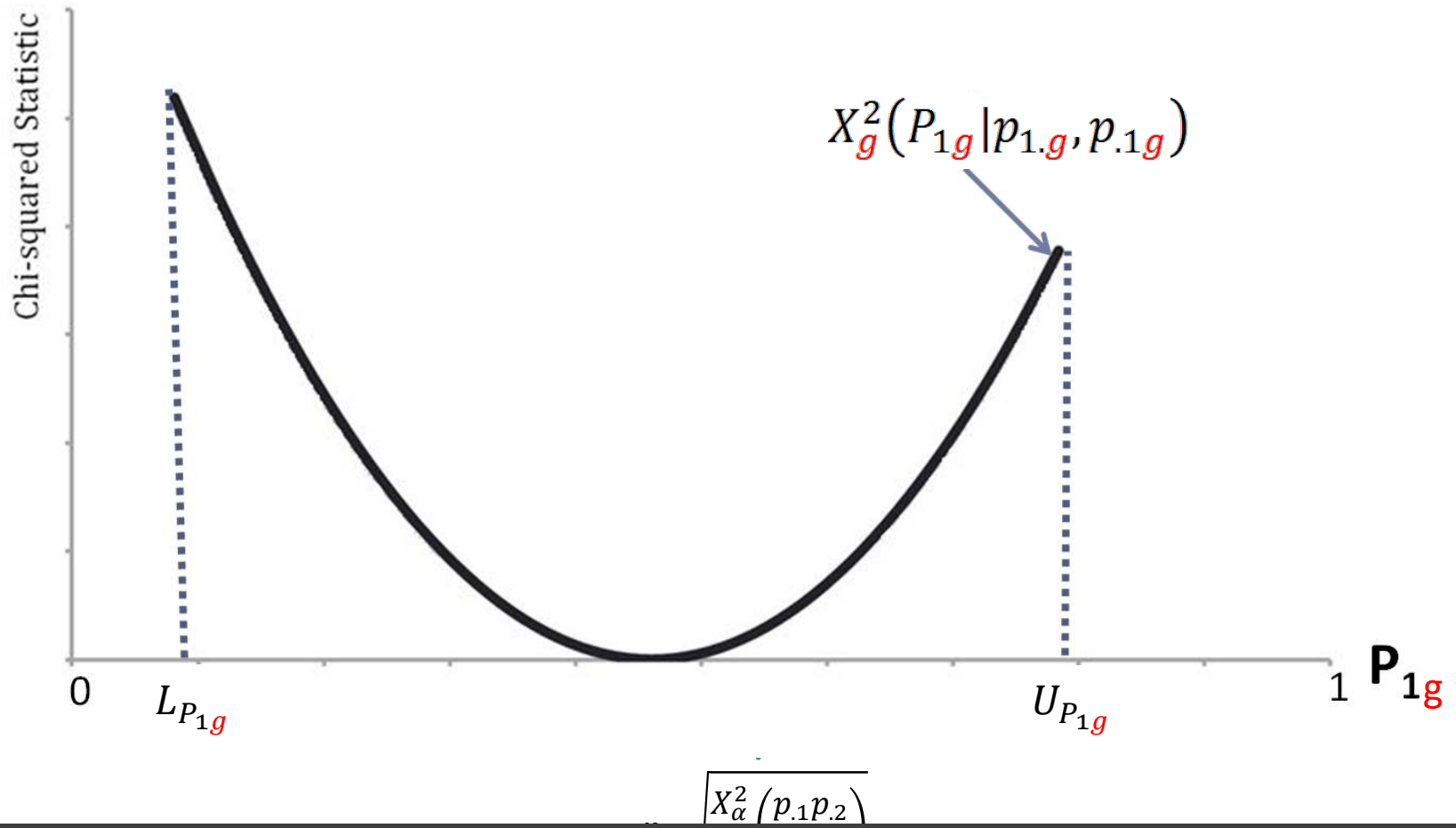
Reformulate the chi-squared statistic

$$X^2 = \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

$$X_g^2(P_{1g}|p_{1.g}, p_{.1g}) = n_g \left(\frac{P_{1g} - p_{.1g}}{p_{2.g}}\right)^2 \left(\frac{p_{1.g}p_{2.g}}{p_{.1g}p_{.2g}}\right)$$

$X_g^2(P_{1g}|p_{1.g}, p_{.1g})$ compared with $\chi_\alpha^2$ ➡ **DECISION**

# AAI curve for each electorate



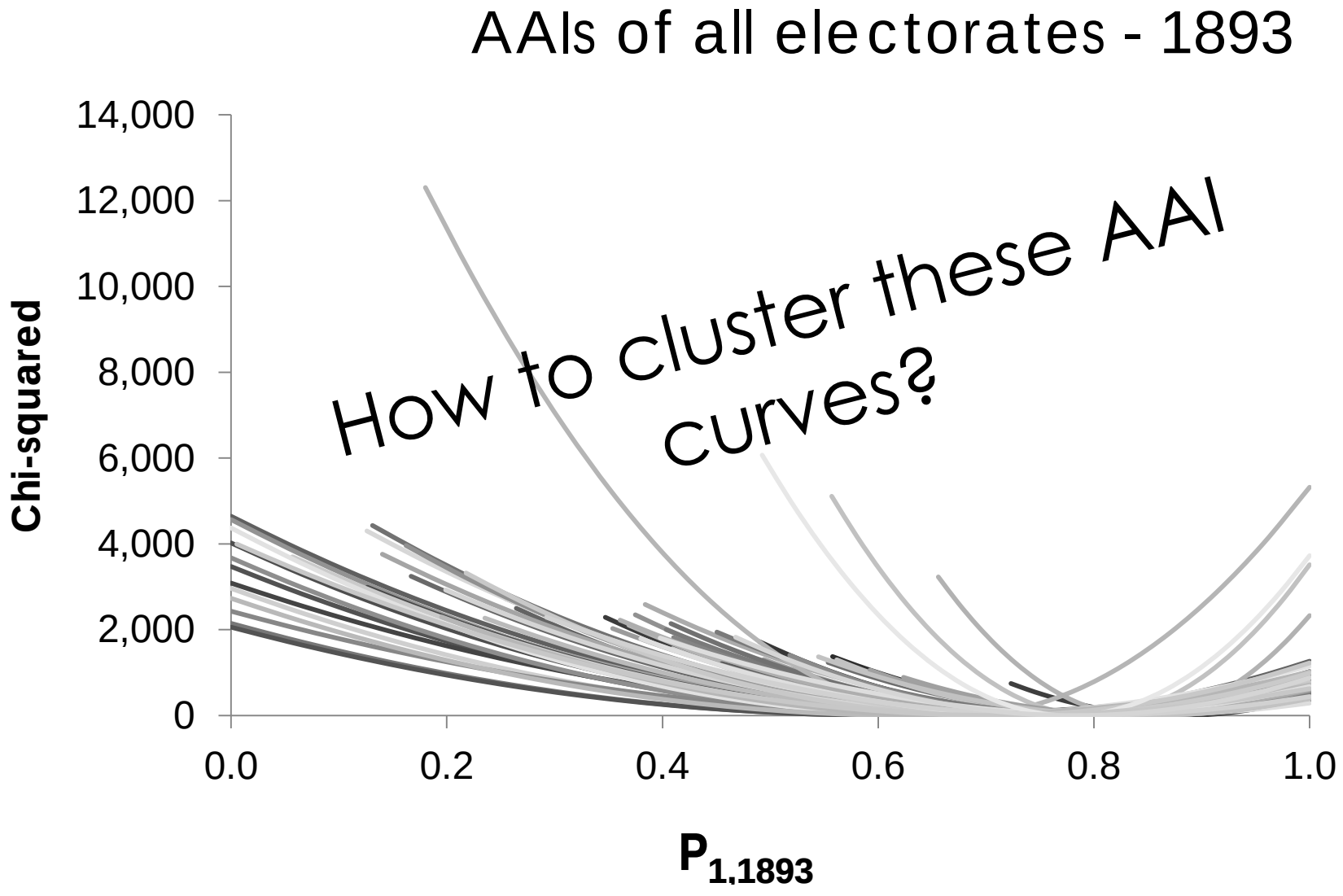

$$A_{\alpha g} = 100\left(1 - \frac{\chi_\alpha^2\left[\left(L_{P_1g}^\alpha - L_{P_1g}\right) + \left(U_{P_1g} - U_{P_1g}^\alpha\right)\right]}{kn_g\left[\left(U_{P_1g} - p_{.1g}\right)^3 + \left(L_{P_1g} - p_{.1g}\right)^3\right]} - \frac{\left(U_{P_1g}^\alpha - p_{.1g}\right)^3 - \left(L_{P_1g}^\alpha - p_{.1g}\right)^3}{\left(U_{P_1g} - p_{.1g}\right)^3 - \left(L_{P_1g} - p_{.1g}\right)^3}\right)$$

# AAI graphical presentation of electorate 1, 1893

| Electorate 1 | Vote | Didn't Vote | Total |
|---|---|---|---|
| Women | | | 1,732 |
| Men | | | 2,589 |
| Total | 3,190 | 1,131 | 4,321 |

$A_{\alpha,1} = 99.37$

- $\alpha = 0.05$
- AAI- = 76.58, AAI+ = 22.79
- $P_{11}$ is $P_1$ of electorate 1


AAIs of all electorates - 1893
How to cluster these AAI curves?
Chi-squared
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
0.0
0.2
0.4
0.6
0.8
1.0
P1.1893

## Parabola in Vertex Form

$y = a(x - h)^2 + k$

## AAI curve of each electorate

$$\chi^2_g\left(P_{1g}\middle|p_{1.g},p_{.1g}\right) = n_g\left(\frac{1}{p_{2.g}}\right)^2\left(\frac{p_{1.g}p_{2.g}}{p_{.1g}p_{.2g}}\right)\left(P_{1g}-p_{.1g}\right)^2$$

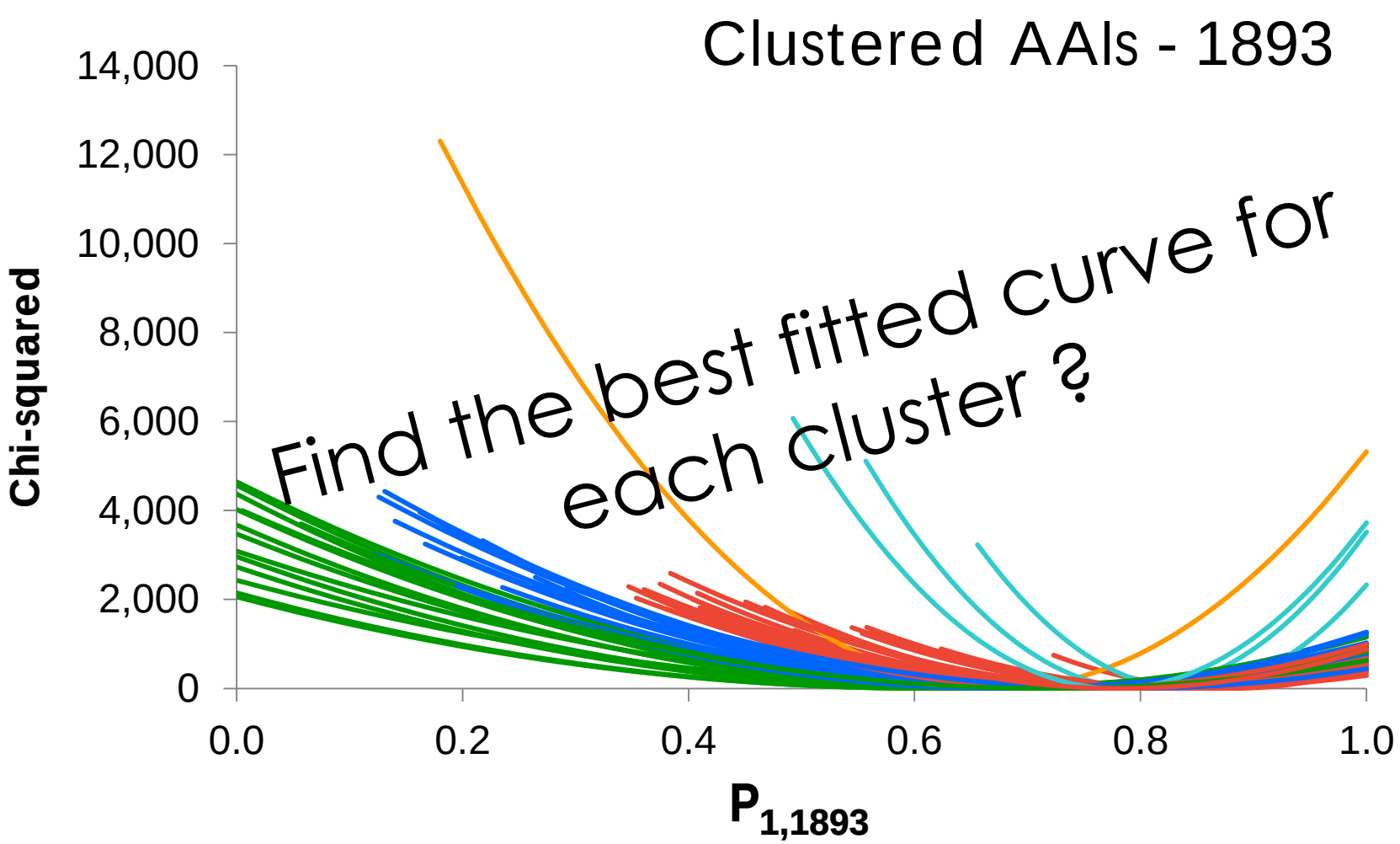
a x h

## Parabola attributes

**curve bounds** $L_{1g} = max\left(0,\frac{n_{.1g}-n_{2.g}}{n_{1.g}}\right) \le P_{1g} \le min\left(\frac{n_{.1g}}{n_{1.g}},1\right) = U_{1g}$

**focus point** $\left(h,k+\frac{1}{4a}\right) \sim \left(p_{.1g},0+\frac{p_{.1g}p_{.2g}p_{2.g}}{4n_g p_{1.g}}\right)$

**vertex point** $(h,k) \sim (p_{.1g},0)$




- Clustering :
  - curve bounds, focus & vertex
  - Mclust

# Calculus

$$y = a\,(x - h)^2 + k$$

$$a = \frac{1}{4d}$$ *d: distance between focus & vertex*

**(h,k)** are the coordinates of the vertex

***Vertex & Focus***

## Optimal parameters for each cluster

**curve bounds** $L_1 = max\left(0, \frac{\sum_{g=1}^{G} n_{.1g} - \sum_{g=1}^{G} n_{2.g}}{\sum_{g=1}^{G} n_{1.g}}\right) \leq P_1 \leq min\left(\frac{\sum_{g=1}^{G} n_{.1g}}{\sum_{g=1}^{G} n_{1.g}}, 1\right) = U_1$

**average focus point** $\left(\frac{1}{G}\sum_{g=1}^{G} p_{.1g}, \frac{1}{G}\sum_{g=1}^{G} \frac{p_{.1g}p_{.2g}p_{2.g}}{4n_g p_{1.g}}\right)$

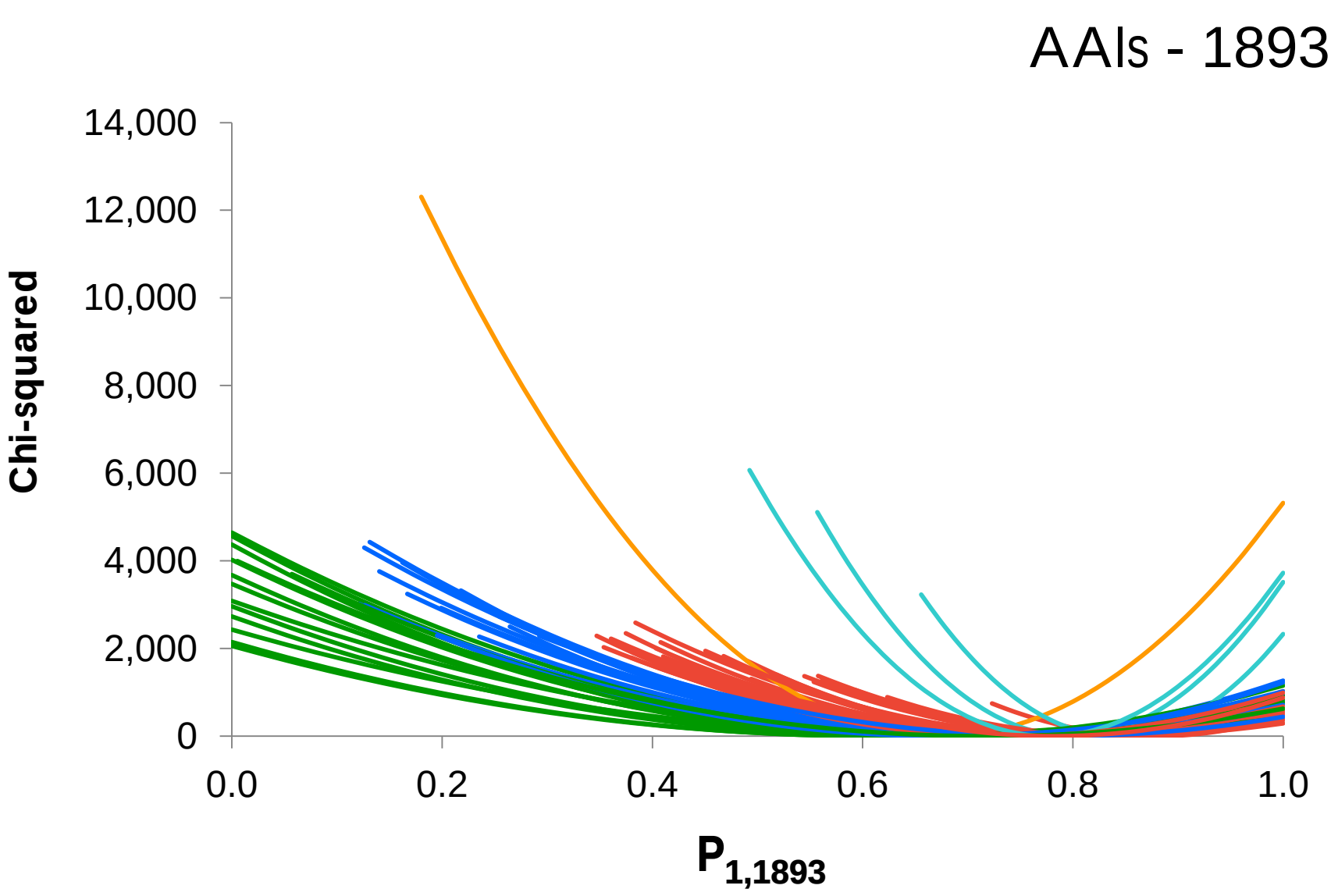
**average vertex point** $\left(\frac{1}{G}\sum_{g=1}^{G} p_{.1g}, 0\right)$

where: $a = \frac{1}{4d}$ *d : distance [avg. Focus; avg. Vertex]*

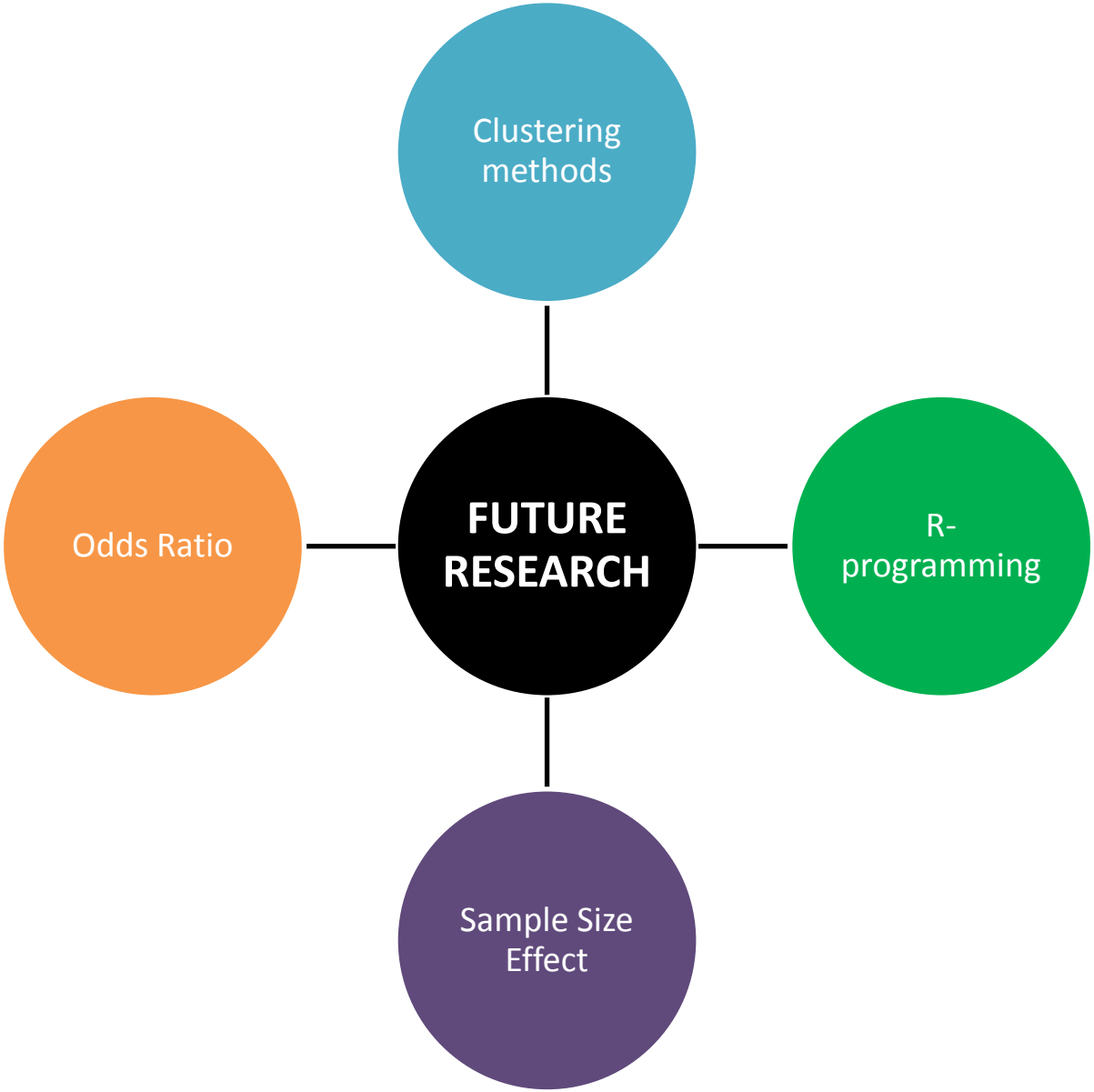
**(h,k)** : *the coordinates of the average vertex*

Fitted average AAI curve

**$y = a\,(x - h)^2 + k$**


AAIs - 1893
Chi-squared
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
0.0
0.2
0.4
0.6
0.8
1.0
P1,1893


Kaitaia
Kaikohe
Kawakawa
Hikurangi
Whangarei
Dargaville
Auckland
Pukekohe
Waiuku
Thames
Whangamata
Waihi
Huntly
Tauranga
Bay of Plenty
Hamilton
North Island
(Te Ika a Maui)
Cambridge
Te Puke
Kawerau
Whakatane
Otorohanga
Tokoroa
Rotorua
Te Kuiti
Murupara
Gisborne
Taupo
Taumarunui
Wairoa
New Plymouth
Inglewood
Stratford
Hawera
Napier
Hastings
Taihape
Patea
Whanganui
Marton
Dannevirke
COOK STRAIT
Palmerston North
Levin
Pahiatua
Waikanae
Masterton
Carterton
Wellington
Motueka
Nelson
Richmond
Blenheim
Seddonville
Westport
Reefton
Rapahoe
Greymouth
Kaikoura
Waiau
Hokitika
Ross
South Island
(Te Wai Pounamu)
Whataroa
Rangiora
Christchurch
Mount Cook
3754 m
Methven
Rolleston
Ashburton
Lake Tekapo
Timaru
Waimate
Kurow
Wanaka
Cromwell
Oamaru
Alexandra
Kingston
Te Anau
Palmerston
Dunedin
Orawia
Gore
Mataura
Otautau
Balclutha
Riverton
Invercargill
Bluff
Stewart Island
NEW ZEALAND
National capital
Region capital
Town
Airport
State Highway
Main road
Railroad
0 50 100 150 km
0 50 100 mi


Clustering methods
Odds Ratio
FUTURE RESEARCH
R-programming
Sample Size Effect

# Reference


[1] E.J. Beh, **"Correspondence analysis of aggregate data: The 2x2 table"**, *Journal of Statistical Planning and Inference*, 138, 2941-2952, 2008.

[2] E.J. Beh, **"The aggregate association index"**, *Computational Statistics and Data Analysis*, 54, 1570 - 1580, 2010.

[3] I.L. Hudson, L. Moore, E.J. Beh, D.G. Steel, **"Ecological inference techniques: an empirical evaluation using data describing gender and voter turnout at New Zealand elections, 1893 – 1919"**, *Journal of the Royal Statistical Society, Series A* 173, 185-213, 2010.

[4]L. Moore, **Was gender a factor in voter participation at New Zealand elections?**, in *Class, Gender and the Vote* (eds M. Fairburn, E. Olssen), Otago University Press: Otago, NZ, pp 129-142, 2005.

thank **you**