NCRM NATIONAL CENTRE FOR RESEARCH METHODS

UKRI Economic and Social Research Council

UNIVERSITY OF Southampton

MANCHESTER 1824 The University of Manchester

THE UNIVERSITY OF EDINBURGH

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Count Data

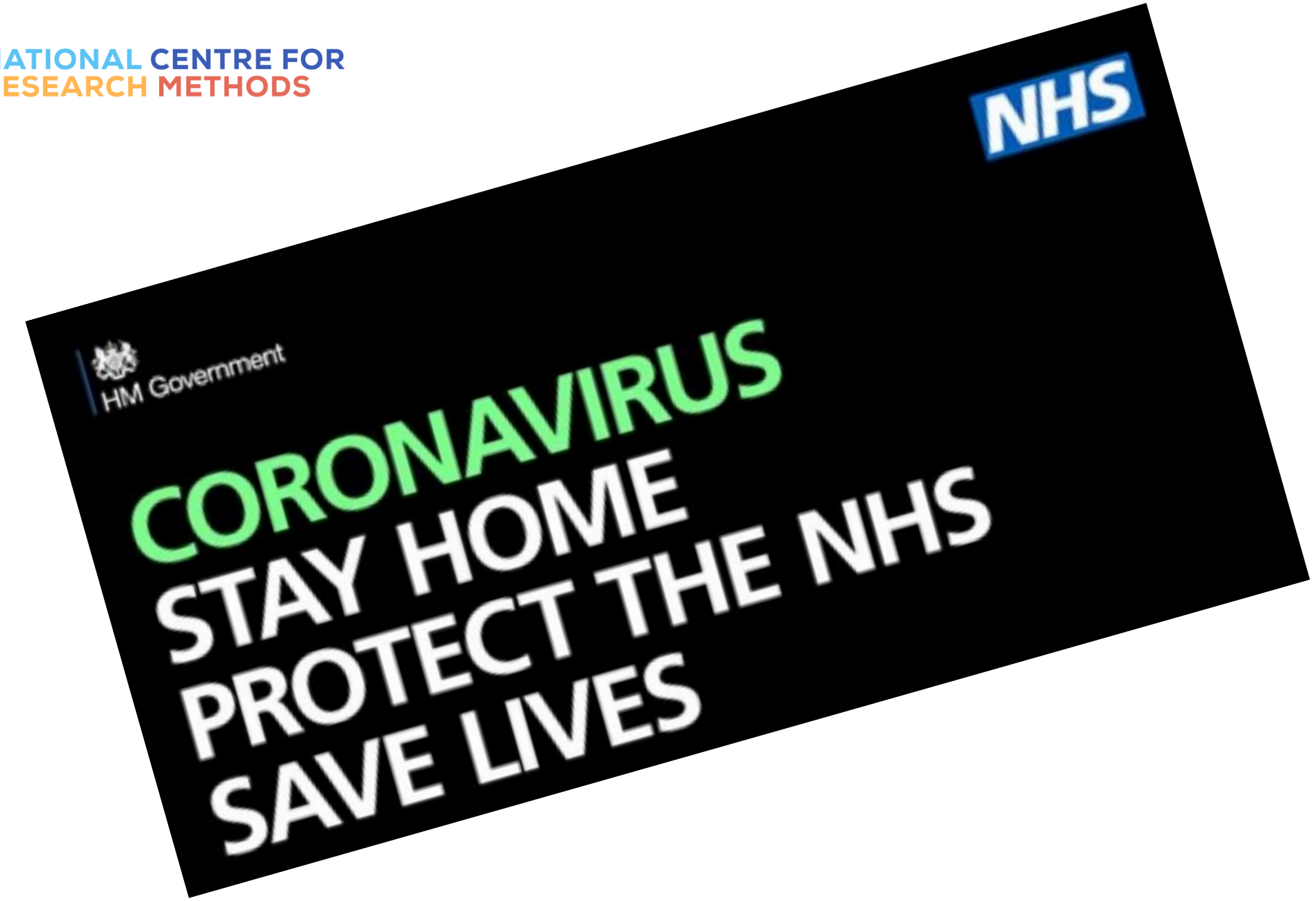
A tale of Poisson and predicting football results

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Professor Vernon Gayle vernon.gayle@ed.ac.uk

@Profbigvern https://github.com/vernongayle

NCRM NATIONAL CENTRE FOR RESEARCH METHODS



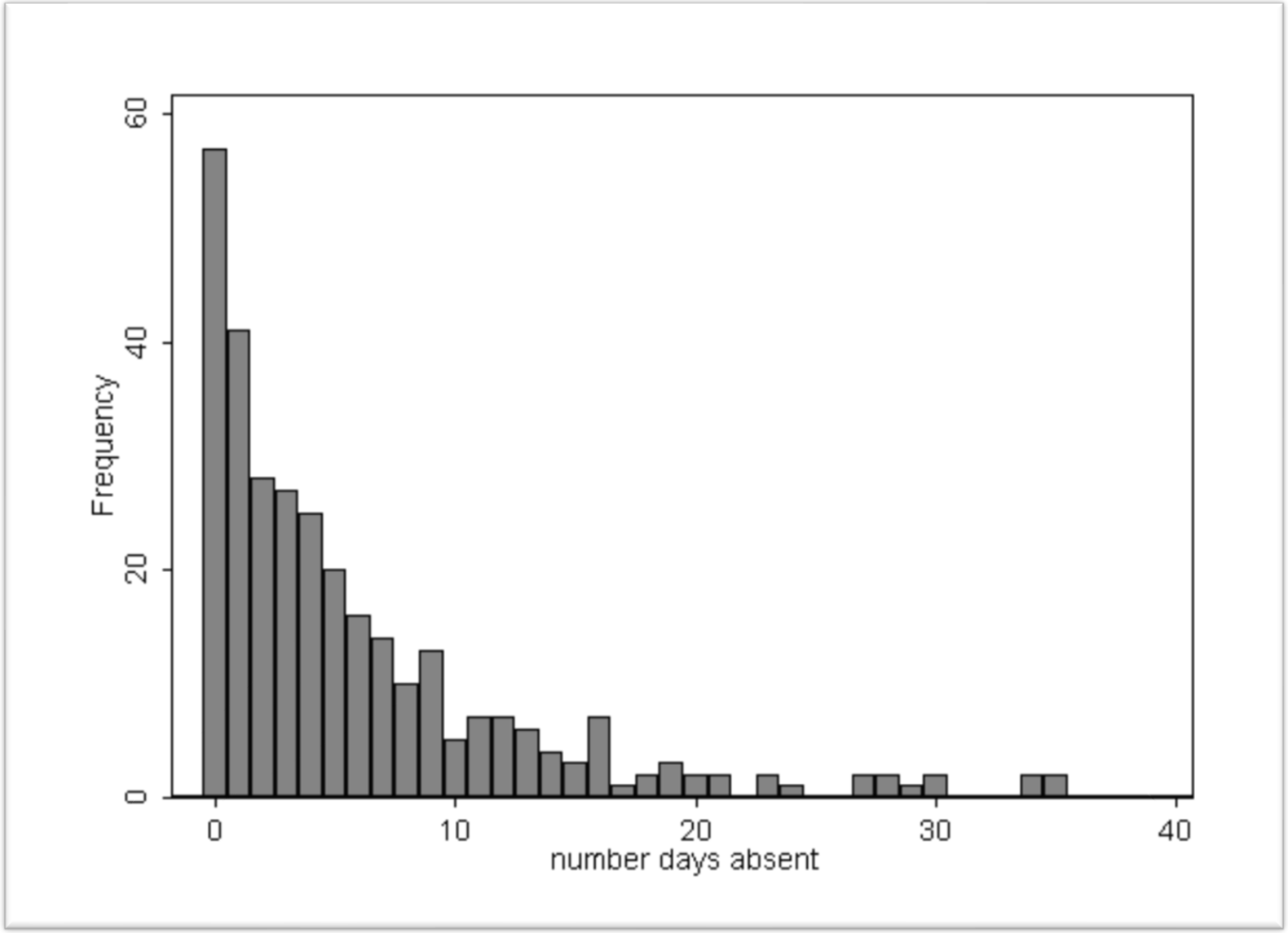
NATIONAL CENTRE FOR RESEARCH METHODS

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Count Data

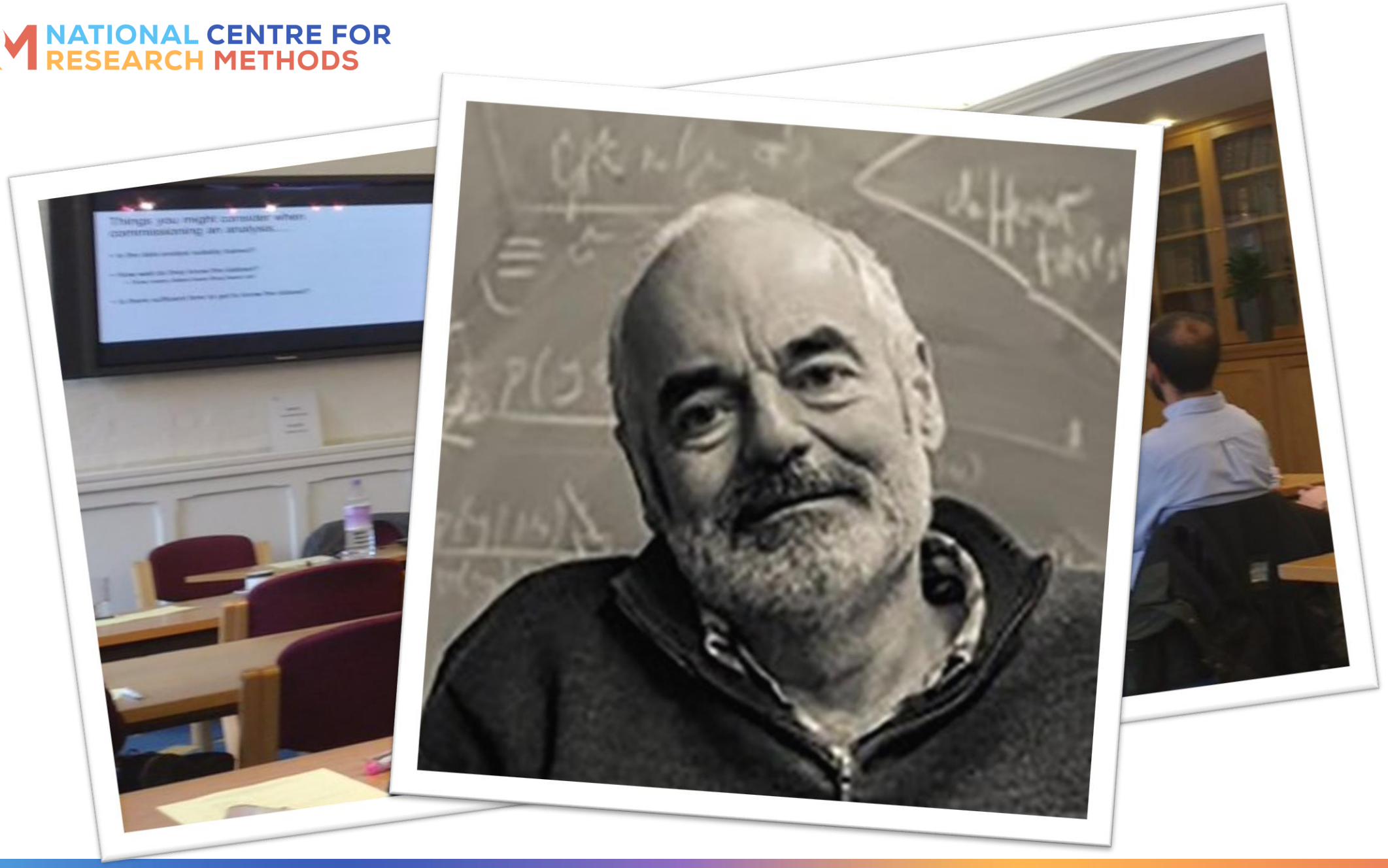
- How many times did you go to the cinema last year?
- How many people has your best friend slept with?
- How many goals have your favourite football team scored this season?


Frequency
0
20
40
60
0
10
20
30
40
number days absent

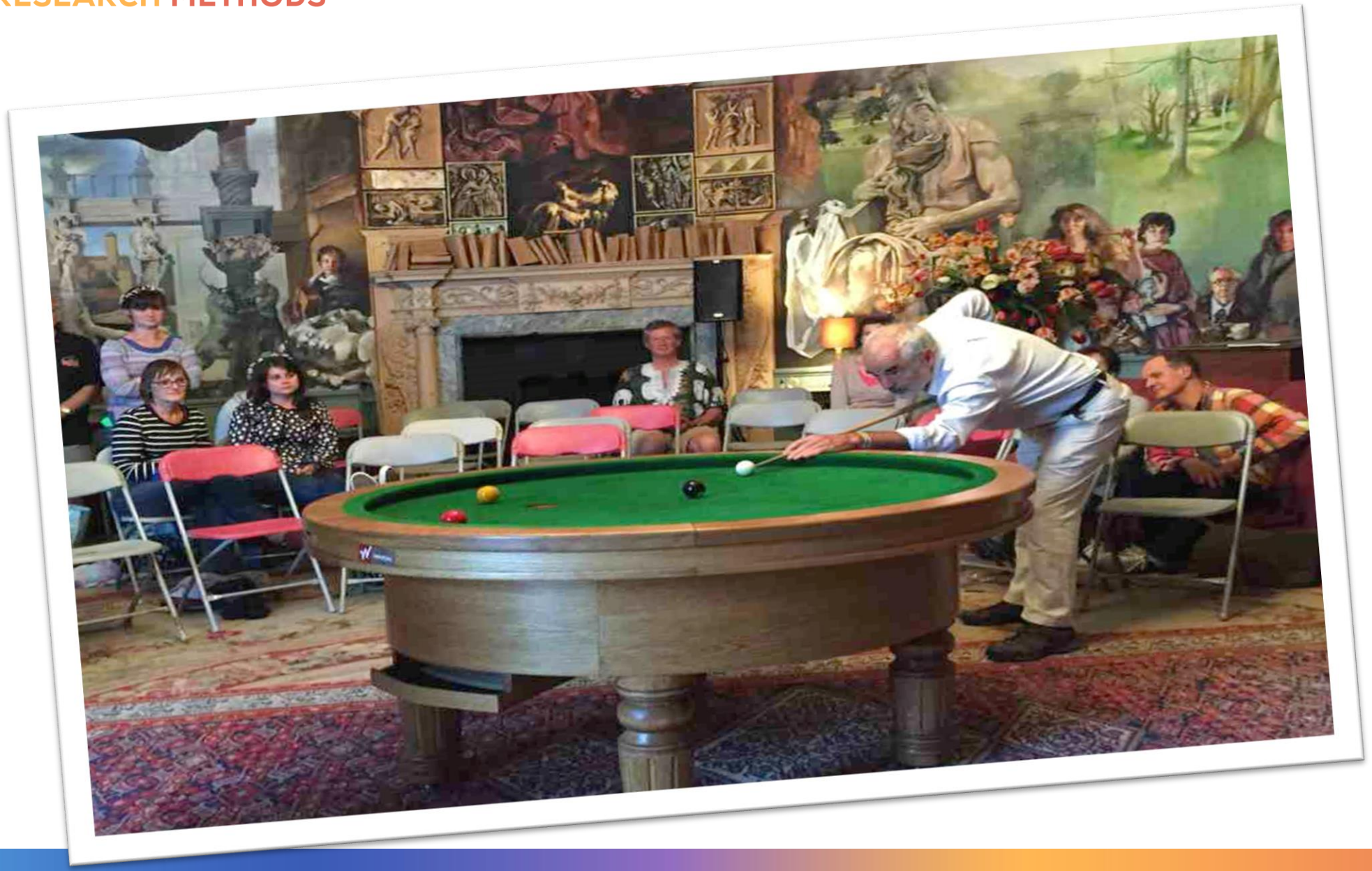
https://en.wikipedia.org/wiki/Sim%C3%A9on_Denis_Poisson

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

NCRM NATIONAL CENTRE FOR RESEARCH METHODS



NCRM NATIONAL CENTRE FOR RESEARCH METHODS



NCRM NATIONAL CENTRE FOR RESEARCH METHODS

NATIONAL CENTRE FOR RESEARCH METHODS

Home Team		Away Team
Clyde	v	Berwick
Cowdenbeath	v	Annan Athletic
Montrose	v	Elgin City
Peterhead	v	Edinburgh City
Stirling Albion	v	Stenhousemuir

Predictions

Home Team		Away Team
Clyde	v	Berwick
Cowdenbeath	v	Annan Athletic
Montrose	v	Elgin City
Peterhead	v	Edinburgh City
Stirling Albion	v	Stenhousemuir

			Predictions
			Fan (i)
Home Team		**Away Team**	
Clyde	v	Berwick	(2-1)
Cowdenbeath	v	Annan Athletic	(0-0)
Montrose	v	Elgin City	(2-0)
Peterhead	v	Edinburgh City	(3-0)
Stirling Albion	v	Stenhousemuir	(2-2)

			Predictions	
			Fan (i)	Dice (ii)
Home Team		**Away Team**		
Clyde	v	Berwick	(2-1)	(2-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)
Montrose	v	Elgin City	(2-0)	(1-4)
Peterhead	v	Edinburgh City	(3-0)	(3-1)
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Team	Played	Won	Drawn	Lost	Goals For	Goals Against	Points
Montrose	35	23	7	5	59	34	76
Peterhead	35	23	4	8	77	38	73
Stirling Albion	35	16	6	13	60	51	54
Stenhousemuir	35	15	8	12	55	46	53
Clyde	35	14	9	12	51	48	51
Elgin City	35	14	6	15	53	60	48
Annan Athletic	35	11	11	13	47	41	44
Berwick	35	8	10	17	29	58	34
Edinburgh City	35	7	9	19	36	60	30
Cowdenbeath	35	4	10	21	23	54	22

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

NATIONAL CENTRE FOR RESEARCH METHODS

Team	Played	Won	Drawn	Lost	Goals For	Goals Against	Points
Montrose	35	23	7	5	59	34	76
Peterhead	35	23	4	8	77	38	73
Stirling Albion	35	16	6	13	60	51	54
Stenhousemuir	35	15	8	12	55	46	53
Clyde	35	14	9	12	51	48	51
Elgin City	35	14	6	15	53	60	48
Annan Athletic	35	11	11	13	47	41	44
Berwick	35	8	10	17	29	58	34
Edinburgh City	35	7	9	19	36	60	30
Cowdenbeath	35	4	10	21	23	54	22

NATIONAL CENTRE FOR RESEARCH METHODS

Team	Played	Won	Drawn	Lost	Goals For	Goals Against	Points
Montrose	35	23	7	5	59	34	76
Peterhead	35	23	4	8	77	38	73
Stirling Albion	35	16	6	13	60	51	54
Stenhousemuir	35	15	8	12	55	46	53
Clyde	35	14	9	12	51	48	51
Elgin City	35	14	6	15	53	60	48
Annan Athletic	35	11	11	13	47	41	44
Berwick	35	8	10	17	29	58	34
Edinburgh City	35	7	9	19	36	60	30
Cowdenbeath	35	4	10	21	23	54	22

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Attack Strength –

how good is the team at scoring goals

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Team	Played	Won	Drawn	Lost	Goals For	Goals Against	Points
Montrose	35	23	7	5	59	34	76
Peterhead	35	23	4	8	77	38	73
Stirling Albion	35	16	6	13	60	51	54
Stenhousemuir	35	15	8	12	55	46	53
Clyde	35	14	9	12	51	48	51
Elgin City	35	14	6	15	53	60	48
Annan Athletic	35	11	11	13	47	41	44
Berwick	35	8	10	17	29	58	34
Edinburgh City	35	7	9	19	36	60	30
Cowdenbeath	35	4	10	21	23	54	22

The average number of goals scored by each team in the league is 49

(i.e. the 10 teams have scored 490 goals in total)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Team	Played	Won	Drawn	Lost	Goals For	Goals Against	Points
Montrose	35	23	7	5	59	34	76
Peterhead	35	23	4	8	77	38	73
Stirling Albion	35	16	6	13	60	51	54
Stenhousemuir	35	15	8	12	55	46	53
Clyde	35	14	9	12	51	48	51
Elgin City	35	14	6	15	53	60	48
Annan Athletic	35	11	11	13	47	41	44
Berwick	35	8	10	17	29	58	34
Edinburgh City	35	7	9	19	36	60	30
Cowdenbeath	35	4	10	21	23	54	22

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Attack Strength –
how good is the team at scoring goals

Stirling Albion 60/49 = 1.22

Stenhousemuir 55/49 = 1.12

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Defensive Weakness –

how bad is the team is at defending (measured by conceding goals)

The average number of goals conceded by each team in the league is 49

(i.e. the 10 teams have conceded 490 goals in total)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Defensive Weakness –
how bad is the team is at defending (measured by conceding goals)

Stirling Albion 51/49 = 1.04

Stenhousemuir 46/49 = 0.94

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Two more measures…

Home Average *–*

The average number of goals home teams score

Away Average *–*

The average number of goals away teams score

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Team	Points	Goals For Home	Games Played Home	Goals For Away	Games Played Away
Montrose	76	30	17	29	18
Peterhead	73	35	17	42	18
Stirling Albion	54	29	17	31	18
Stenhousemuir	53	32	18	23	17
Clyde	51	22	17	29	17
Elgin City	48	31	18	22	17
Annan Athletic	44	24	18	23	17
Berwick	34	21	18	8	17
Edinburgh City	30	19	18	17	18
Cowdenbeath	22	12	17	11	18
Totals		*255*	*175*	*235*	*175*

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Home Average **–**

The average number of goals home teams score

The average number of goals that home teams score in a league game is 1.46 (255/175)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Team	Points	Goals For Home	Games Played Home	Goals For Away	Games Played Away
Montrose	76	30	17	29	18
Peterhead	73	35	17	42	18
Stirling Albion	54	29	17	31	18
Stenhousemuir	53	32	18	23	17
Clyde	51	22	17	29	17
Elgin City	48	31	18	22	17
Annan Athletic	44	24	18	23	17
Berwick	34	21	18	8	17
Edinburgh City	30	19	18	17	18
Cowdenbeath	22	12	17	11	18
Totals		*255*	*175*	*235*	*175*

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Away Average –

The average number of goals away teams score

The average number of goals that away teams score in a league game is 1.34 (235/175)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Expect Goals

Stirling Albion v Stenhousemuir

NATIONAL CENTRE FOR RESEARCH METHODS

Home Team

Average Goals per match home	Attack Strength	Defensive Weakness	Expected Goals
1.46	1.22	0.94	1.67

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Away Team

Average Goals per match away	Attack Strength	Defensive Weakness	Expected Goals
1.34	1.12	1.04	1.56

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

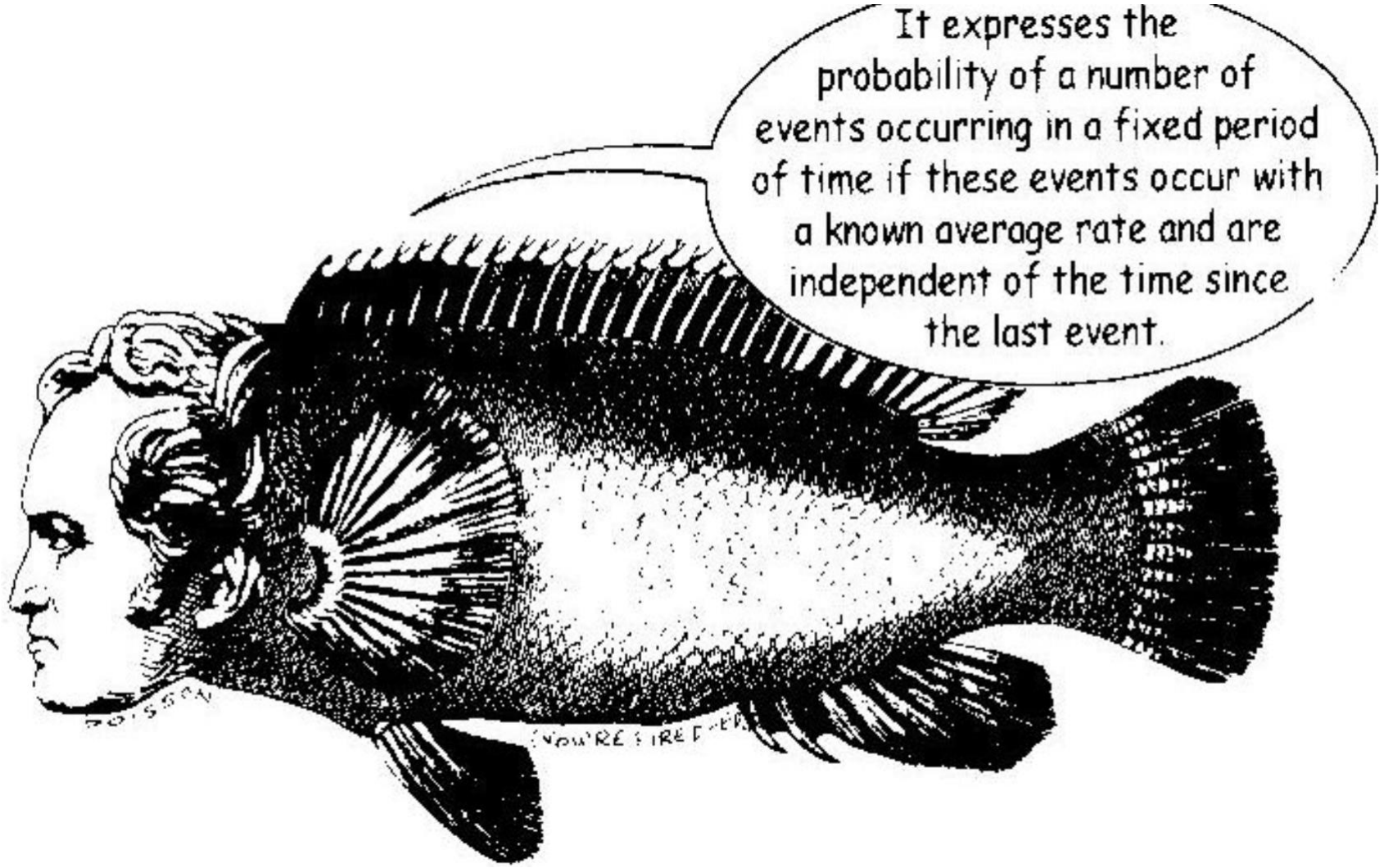
Expected Goals

Average Goals per match home	Attack Strength	Defensive Weakness	Expected Goals
1.46	1.22	0.94	**1.67**

Average Goals per match away	Attack Strength	Defensive Weakness	Expected Goals
1.34	1.12	1.04	**1.56**

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

The Poisson Distribution & Expected Goals



https://fluffypool.wordpress.com/tag/poisson-distribution/

$$Pr = e^{-\lambda} (\lambda^k / k!)$$

$$Pr = e^{-\lambda} (\lambda^k / k!)$$

λ is the expected number of goals

e = 2.71828

(this Euler's number which is a mathematical constant)

$$Pr = e^{-\lambda} (\lambda^k / k!)$$

k is the number of events
(in this example 0 through to 6 goals)

k! is k factorial

$$Pr = e^{-\lambda} (\lambda^k / k!)$$

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Plugging the information for Stirling Albion into this formula…

For one goal the probability is

$$Pr = 2.71828^{-1.67} (1.67^{1} / (1)) = 0.31$$

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Plugging the information for Stirling Albion into this formula…

For two goals the probability is

$$Pr = 2.71828^{-1.67} (1.67^2 / (1 \times 2)) = 0.26$$

	Number of Goals Predicted						
	0	1	2	3	4	5	6
Stirling Albion	0.19	0.31	0.26	0.15	0.06	0.02	0.01

	Number of Goals Predicted						
	0	1	2	3	4	5	6
Stirling Albion	0.19	0.31	0.26	0.15	0.06	0.02	0.01
Stenhousemuir	0.21	0.33	0.26	0.13	0.05	0.02	0.00

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

	Number of Goals Predicted						
	0	1	2	3	4	5	6
Stirling Albion	0.19	0.31	0.26	0.15	0.06	0.02	0.01
Stenhousemuir	0.21	0.33	0.26	0.13	0.05	0.02	0.00

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	
Montrose	v	Elgin City	(2-0)	(1-4)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	
Montrose	v	Elgin City	(2-0)	(1-4)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)
Montrose	v	Elgin City	(2-0)	(1-4)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions		
			Fan (i)	Dice (ii)	Statistical Method (iii)
Home Team		**Away Team**			
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)

jupyter ncrm_poisson_20200830_vg_v1 Last Checkpoint: 21 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import poisson
```

In [2]:
```
x1= np.arange(0,7,1)
home_goals = (x1, poisson.pmf(x1,1.67))
print(home_goals)
```

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

More Complex Models

- Emphasis on recent results
- Different home advantage than the league average
- The composition of individual teams (e.g. new players; injuries etc.)
- Weather
- State of the pitch
- Expert knowledge


SATURDAY 28TH APRIL 2018

			Predictions			
			Fan	Dice	Statistical Method	Result
			(i)	(ii)	(iii)	(iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions			
			Fan	Dice	Statistical Method	Result
			(i)	(ii)	(iii)	(iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	(1-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	

			Predictions			
			Fan (i)	Dice (ii)	Statistical Method (iii)	Result (iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	(1-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	(0-2)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

			Predictions			
			Fan	Dice	Statistical Method	Result
			(i)	(ii)	(iii)	(iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	(1-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	(0-2)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	(1-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	

			Predictions			
			Fan (i)	Dice (ii)	Statistical Method (iii)	Result (iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	(1-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	(0-2)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	(1-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	(2-1)
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	

			Predictions			
			Fan (i)	Dice (ii)	Statistical Method (iii)	Result (iv)
Home Team		**Away Team**				
Clyde	v	Berwick	(2-1)	(2-2)	(1-0)	(1-2)
Cowdenbeath	v	Annan Athletic	(0-0)	(5-3)	(0-1)	(0-2)
Montrose	v	Elgin City	(2-0)	(1-4)	(2-1)	(1-1)
Peterhead	v	Edinburgh City	(3-0)	(3-1)	(2-0)	(2-1)
Stirling Albion	v	Stenhousemuir	(2-2)	(6-5)	(1-1)	(1-1)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

The Outcomes

Stats method	1 / 5 correct scores	3 / 5 correct results
Fan	0 / 5 correct scores	2 / 5 correct results
Dice	0 / 5 correct scores	1 / 5 correct results


CAUTION

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

We **do not** advocate using these methods for gambling

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

More Complex Models

- Simple Poisson models only a few terms

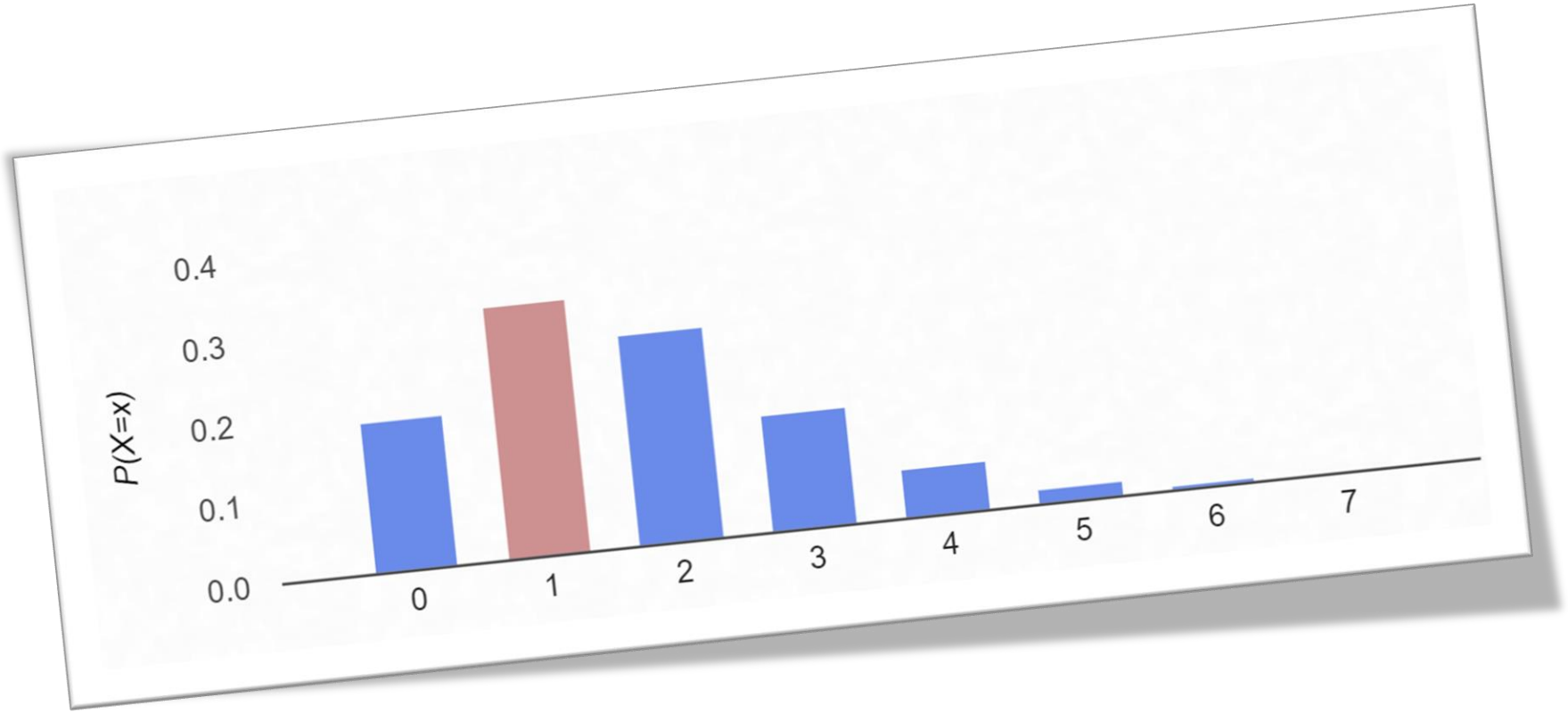
(i.e. home advantage x attack strength x defensive weakness)

- But we could extend these models as I have noted above

- You may have also thought about some additional measures!

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

	0	1	2	3	4	5	6
0							
1							
2							
3							
4							
5							
6							


0.4
0.3
0.2
0.1
0.0
P(X=x)
0
1
2
3
4
5
6
7

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Analysing Social Science Data

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Left Hand Side = Right Hand Side + Error

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

https://stats.idre.ucla.edu/other/dae/

Count Models					
Poisson Regression	Stata	SAS	SPSS	Mplus	R
Negative Binomial Regression	Stata	SAS	SPSS	Mplus	R
Zero-inflated Poisson Regression	Stata	SAS		Mplus	R
Zero-inflated Negative Binomial Regression	Stata	SAS		Mplus	R
Zero-truncated Poisson	Stata	SAS			R
Zero-truncated Negative Binomial	Stata	SAS		Mplus	R

Parental social class and school GCSE outcomes: two decades of evidence from UK household panel surveys

Sarah Stopforth , Vernon Gayle & Ellen Boeren

To cite this article: Sarah Stopforth , Vernon Gayle & Ellen Boeren (2020): Parental social class and school GCSE outcomes: two decades of evidence from UK household panel surveys, Contemporary Social Science, DOI: 10.1080/21582041.2020.1792967

To link to this article: https://doi.org/10.1080/21582041.2020.1792967

Table 2. Zero-inflated negative binomial regression model – number of GCSEs at grades A*–C (BHPS).

				Quasi-Variance		
	Coefficient	Standard Error		Standard Error	Lower Comparison Interval	Upper Comparison Interval
Logistic estimation: Zero A*–Cs						
Parental NS-SEC						
1.1 Large employers and higher managerial occupations	0.64	(0.67)		–	–	–
1.2 Higher professional occupations	Ref.	(.)		–	–	–
2 Lower managerial and professional occupations	0.80	(0.49)		–	–	–
3 Intermediate occupations	0.65	(0.55)		–	–	–
4 Small employers and own account workers	1.43	(0.50)	**	–	–	–
5 Lower supervisory and technical occupations	1.22	(0.55)	*	–	–	–
6 Semi-routine occupations	1.62	(0.53)	**	–	–	–
7 Routine occupations	1.36	(0.54)	*	–	–	–
Parental Education Level						
Higher education	Ref.	(.)		–	–	–
Further education	0.73	(0.39)		–	–	–
School-level education	0.87	(0.41)	*	–	–	–
Below school-level education	1.38	(0.45)	**	–	–	–
Housing Tenure						
Owned or privately rented	Ref.	(.)		–	–	–
Social housing	1.23	(0.23)	***	–	–	–
Gender						
Male	Ref.	(.)		–	–	–
Female	−0.68	(0.17)	***	–	–	–
Constant	−2.67	(0.63)	***			

NATIONAL CENTRE FOR RESEARCH METHODS

How to cite this video

Gayle, V. (2020) Count Data - *A tale of Poisson and predicting football results.* Available at: https://www.ncrm.ac.uk (Accessed: day month year)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Count Data

A tale of Poisson and predicting football results

Professor Vernon Gayle
vernon.gayle@ed.ac.uk
@Profbigvern
https://github.com/vernongayle

UKRI Economic and Social Research Council

University of Southampton

Manchester 1824 The University of Manchester

The University of Edinburgh