
1
2
3

1

"It's easy to lie with statistics, but it's hard to tell the truth without them."

— Andrejs Dunjels

1

What research **question** do I have?

What **argument** do I want to make?

What are the observable **implications** of my argument?

Which descriptive and inferential statistics do I want to use?

1

Neither the <u>mean</u> nor the <u>median</u> is hard to calculate.

What is harder is deciding <u>which one</u> gives a more accurate measure of the middle observation in a particular situation.

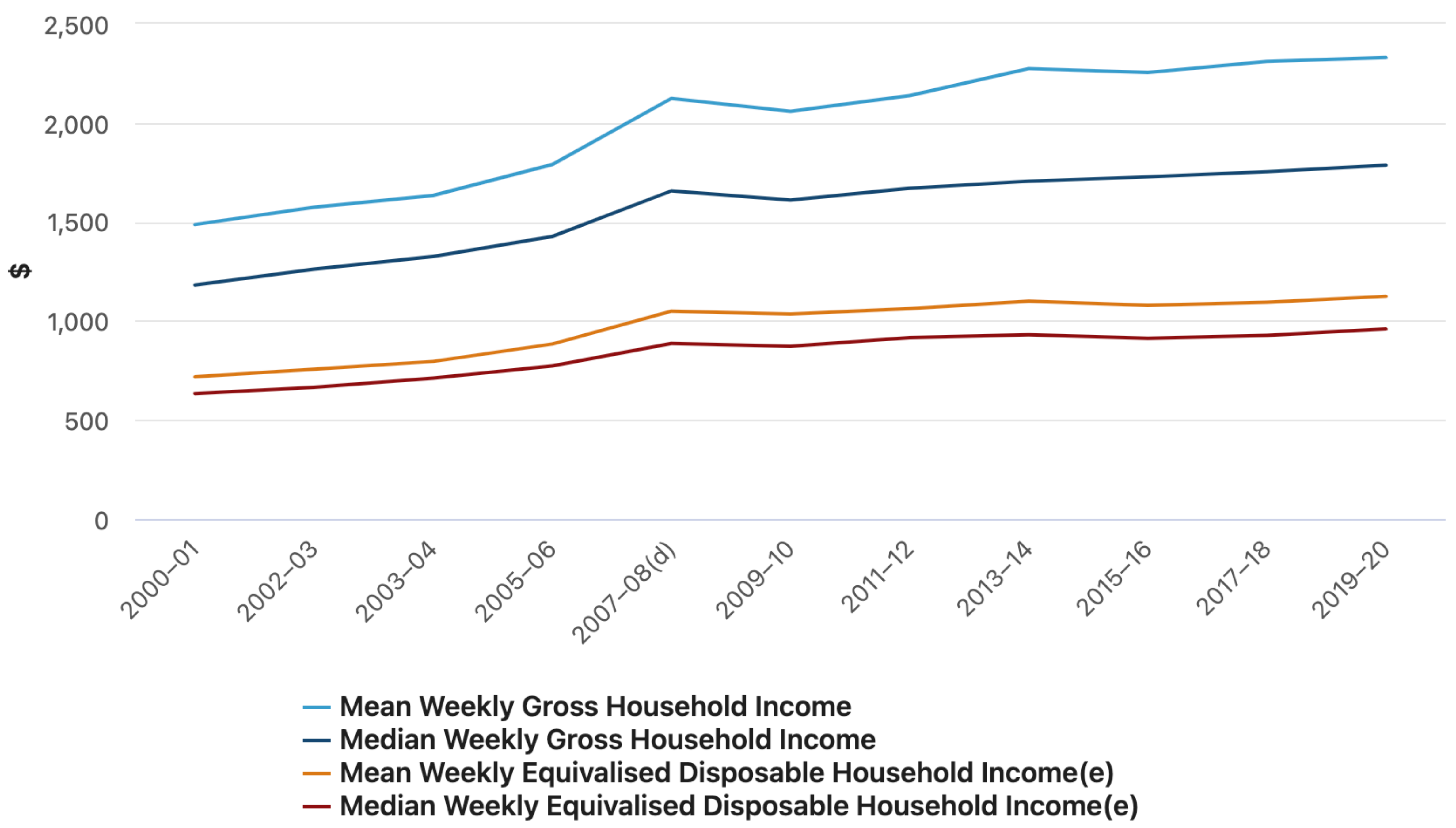
Median values are not sensitive to <u>outliers</u>.

Source: https://www.chicagotribune.com/dining/ct-food-replay-its-always-sunny-in-philadelphia-pop-up-20180228-story.html

Source: https://www.haaretz.com/israel-news/2018-03-19/ty-article/is-tesla-coming-to-israel-elon-musk-appears-in-jerusalem-bar/0000017f-e5d1-dc7e-adff-f5fdd43d0000

1

Graph 1 - Weekly household income, Australia, 2000–01 to 2019–20(a)(b)(c)



a. Survey of Income and Housing data was collected in labelled years
b. In 2019–20 dollars, adjusted using changes in the Consumer Price Index
c. Due to the change in collection methodology, estimates may not be directly comparable to previous cycles. For more information please see the Household Income and Wealth, Australia - Methodology, 2019–20 financial year
d. In 2007–08 there was a change in income standards, see the Household Income and Wealth, Australia - Methodology, 2019–20 financial year for more information
e. Equivalised disposable household income estimates are adjusted by equivalence factors to standardise them for variations in household size and composition, while taking into account the economies of scale that arise from the sharing of dwellings

Source(s): ABS Survey of Income and Housing, various years

mean: $2,329

median: $1,786

Source:https://www.abs.gov.au/statistics/economy/finance/household-income-and-wealth-australia/latest-release

1

How does **basic probabilities** underly most political science research?

Can we better understand probabilities through a few **examples**?


2

2

“Probability is the study of events and outcomes involving an element of **uncertainty**,” (Wheelan 2013: 71).

2

Because most of the time political scientists are dealing with **samples** instead of **populations**.

Probabilities help us determine which relationships are **statistically significant**.

In other words, they are unlikely to occur by **chance**.

2

1. Key probability properties
2. Central limit theorem
3. Standard normal distribution
4. Confidence intervals
5. Expected values

2

1. All outcomes have a **probability** ranging from **0 to 1**.

Sources: Kellstedt & Whitten (2018: 147) and Wheelan (2013: 76)

2

Sources:https://graziano-raulin.com/statistics/concepts/probability.htm

2

$$\text{Probability} = \frac{\text{Number of relevant outcomes}}{\text{Total number of possible outcomes}}$$

Sources:https://graziano-raulin.com/statistics/concepts/probability.htm

2

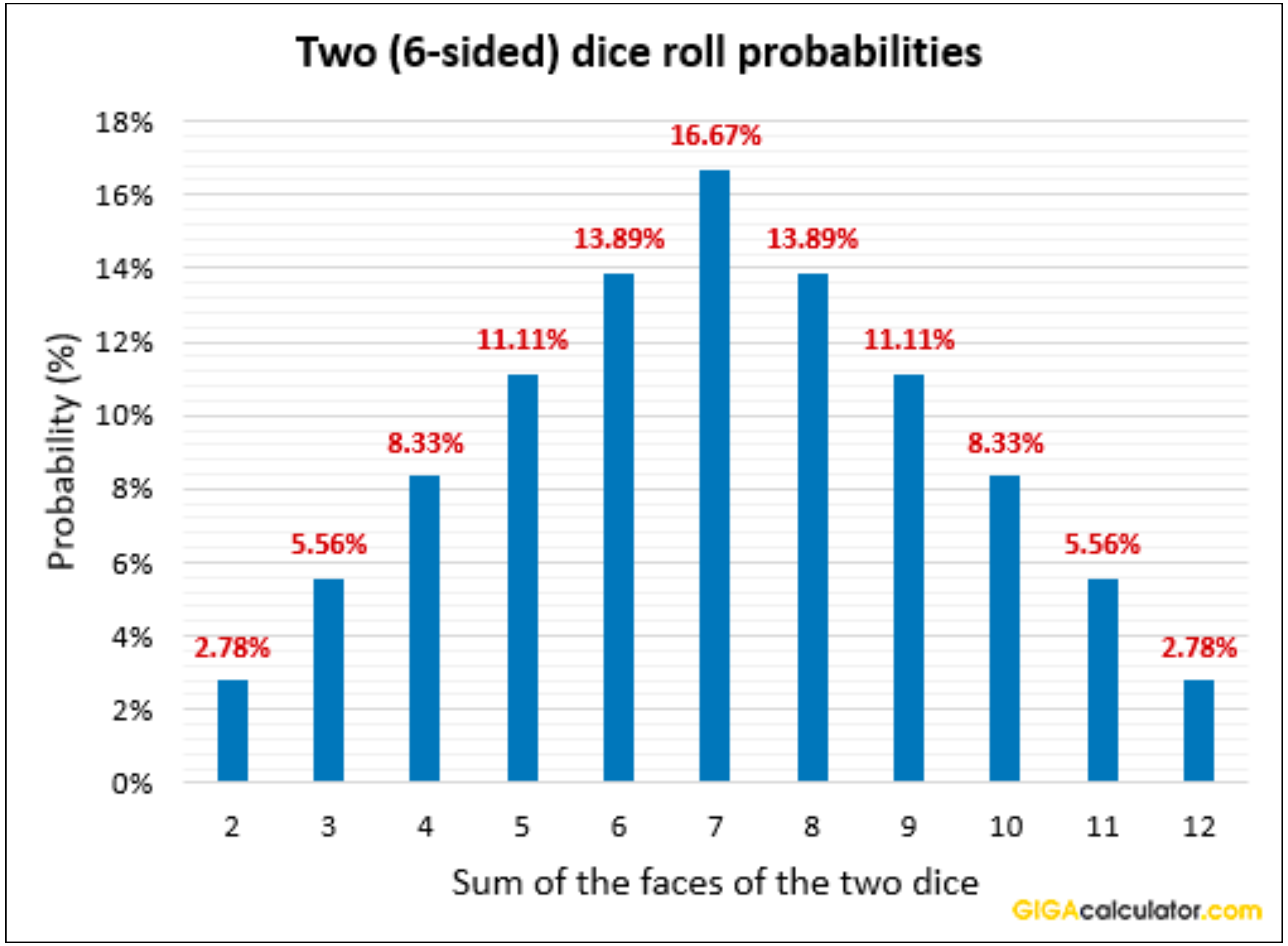
2. The **sum** of all possible outcomes must be exactly **1**.

The **sum** of probabilities of rolling all possible die rolls is:

$$p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1$$

Sources: Kellstedt & Whitten (2018: 147) and Wheelan (2013: 76)

2



Source: https://www.gigacalculator.com/calculators/dice-probability-calculator.php

2

3. If (and only if) two outcomes are **independent**, then the probability of those events both occurring is equal to the **product** of them individually.

So if we want to know the probability of me rolling a **3** and Sajjad rolling a **6**, then we multiply the probability of me rolling a 3 (0.167) and the probability of Sajjad rolling a 6 (0.167) = **0.028**.

Sources: Kellstedt & Whitten (2018: 147) and Wheelan (2013: 76)

2

4. The chance of **either of two outcomes** happening is the **sum** of their probabilities if the options are **mutually exclusive**.

So the chance of **either** rolling a **3** or a **6** is = $p(3) + p(6) = 0.33$.

Sources: Kellstedt & Whitten (2018: 147) and Wheelan (2013: 76)

2

5. If the events are **not mutually exclusive**, the probability of getting A or B consists of the **sum** of their **individual** probabilities **minus** the probability of **both** events happening.

Sources: Kellstedt & Whitten (2018: 147) and Wheelan (2013: 76)

2

1. Assuming events are **independent** when they are not (e.g., rain today and tomorrow).

2. Assuming events are **not independent** when they are (e.g., hot streaks).

3. **Clusters** do happen (e.g., getting struck by lightning).

4. There is often **reversion to the mean** (e.g. doing well on an exam).

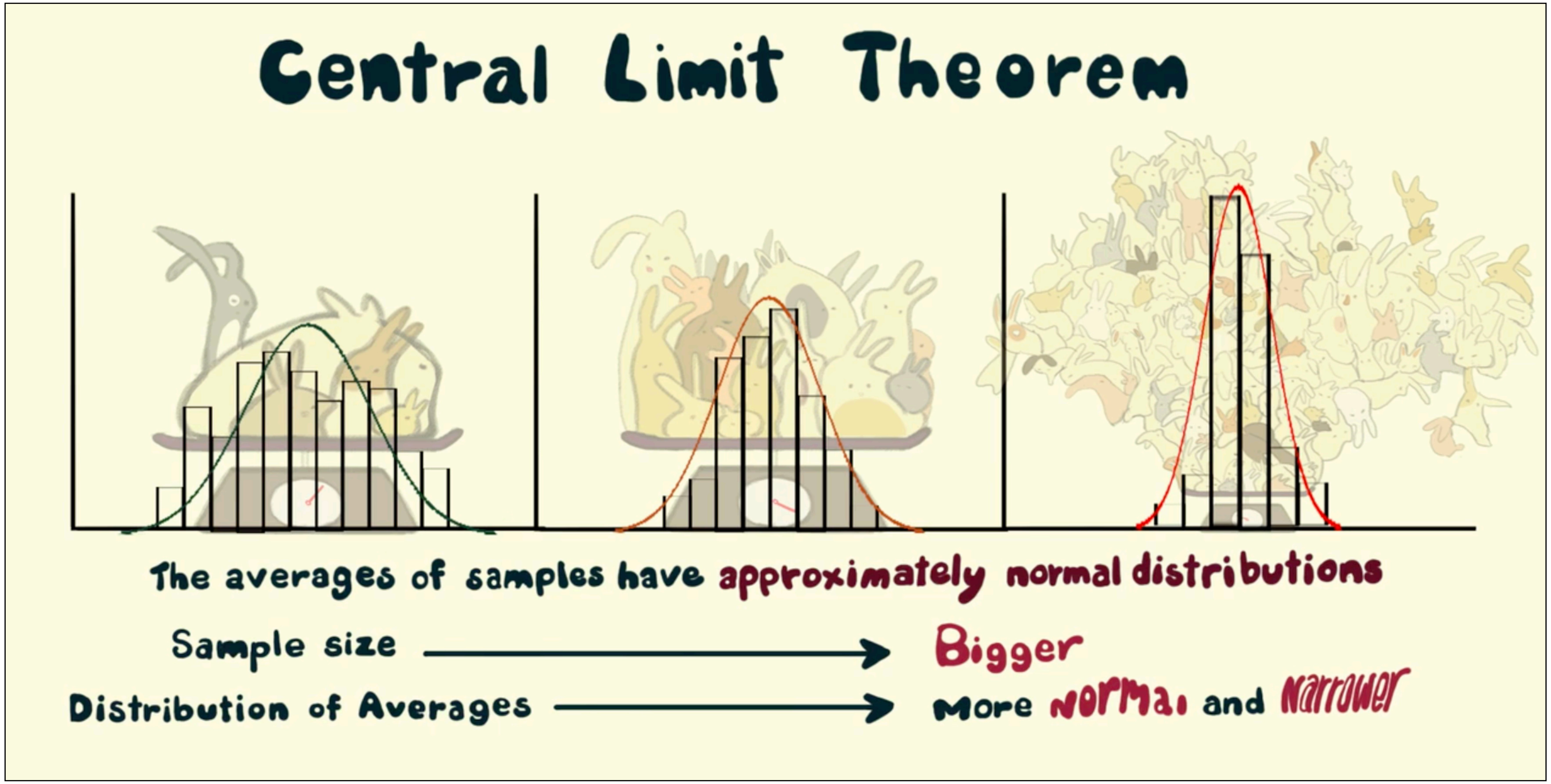
5. Moving from **aggregate statistics to predicting individual** behaviour (e.g., profiling/ecological fallacy).

6. **Garbage** in, garbage out (e.g., data quality).

7. **Analytical tools** are moving faster than our knowledge of what to do with results (e.g. predictive AI, black swans).


2

2



Source: https://miro.medium.com/v2/resize:fit:3796/format:png/1*AhMCbLVd5s82XV6M4KTK6A.png

2

Sample size has to be large (say greater than 30 observations).

The **sample mean** will be distributed roughly as a normal distribution around the **population mean**.

The **sample standard deviation** (a.k.a. standard error) will equal the **population standard deviation** over the square root of the number of sample observations.

2

Key point: The **sampling distribution** is normally shaped even though the underlying **frequency distribution** is **not** normally shaped.

2

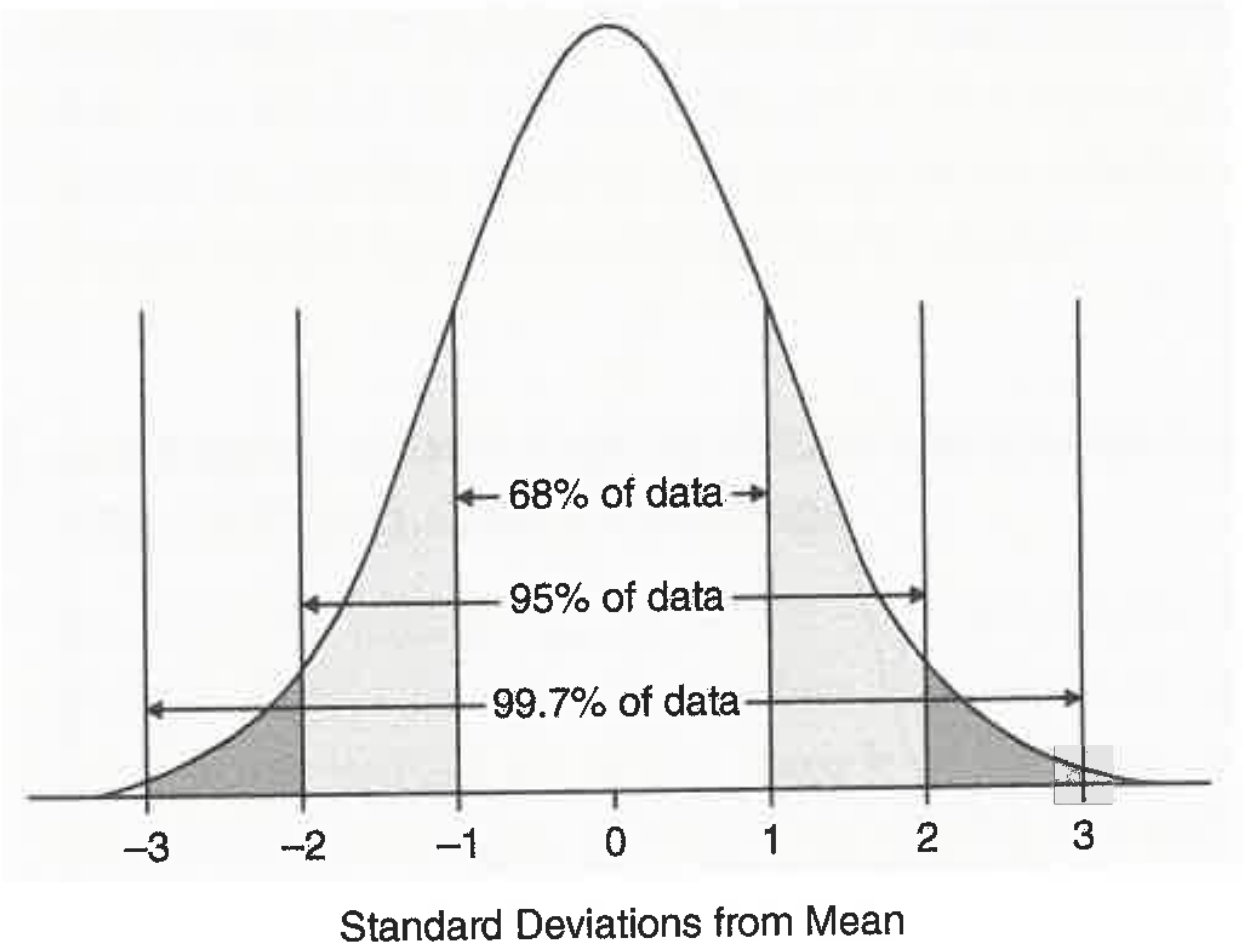


Figure 7.2 The 68–95–99 rule

Source: Kellstedt and Whitten (2018: 150)

2

It is **symmetrical** about the mean

The median, mean, and mode are **the same**.

It has a **predictable area** under the curve within a specific distance of the mean.

Skewness and **kurtosis** are zero.

2

The **standard deviation** is:

$$sd_{\overline{Y}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_1 - \overline{Y})^2}{n-1}}$$

Where:

$sd_{\overline{Y}}$ is the standard deviation of the sample mean.

Y_1 is a value of y for observation 1.

$\overline{Y}$ is the sample mean.

n is the sample size

2

Here is the standard error equation.

$$\sigma_{\overline{Y}} = \frac{sd_Y}{\sqrt{n}}$$

Where:

$\sigma_{\overline{Y}}$ = standard error of the sample mean

sd =standard deviation

n =sample size

This allows us to calculate the confidence intervals around the mean value.

2

A distribution of sample values is the **frequency distribution**, which does not have to be normal.

However, because of the central limit theorem, the repeated sampling **distribution mean will** be naturally distributed, even if the underlying frequency is not.

2

To calculate the confidence interval you need four things

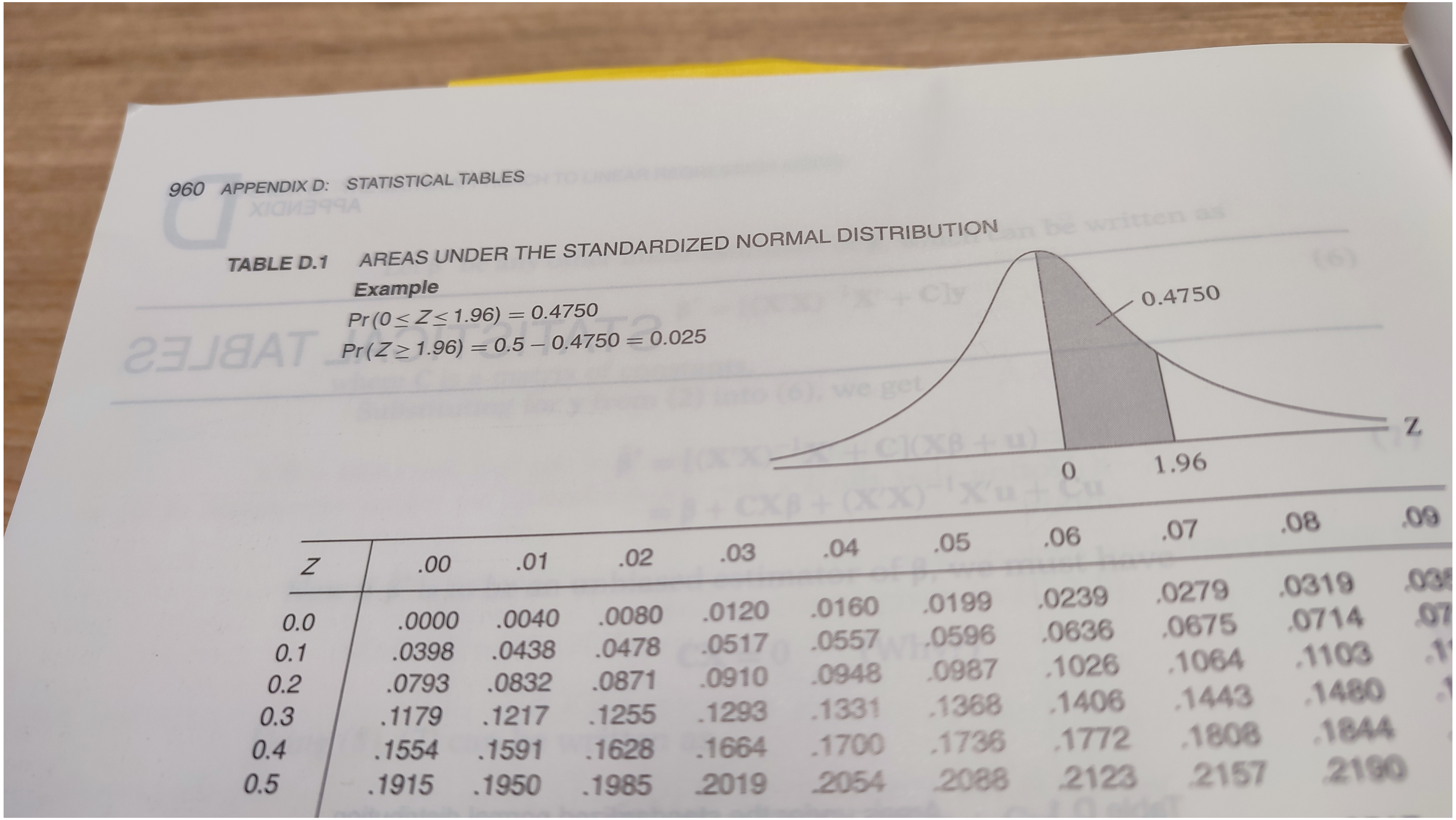
The number of observations (*n*)
The mean ($\bar{X}$)
The standard deviation (s)
The desired confidence level (let's say 95%) you go to the *Z* table and find the *Z*(0.95) score, which is 1.96.

Then you plug these values into the following equation

$$\bar{x} \pm Z\frac{s}{\sqrt{n}}$$

The s/sqrt(n) is the standard error.

2



960 APPENDIX D: STATISTICAL TABLES

TABLE D.1 AREAS UNDER THE STANDARDIZED NORMAL DISTRIBUTION

Example

$Pr(0 \le Z \le 1.96) = 0.4750$

$Pr(Z \ge 1.96) = 0.5 - 0.4750 = 0.025$

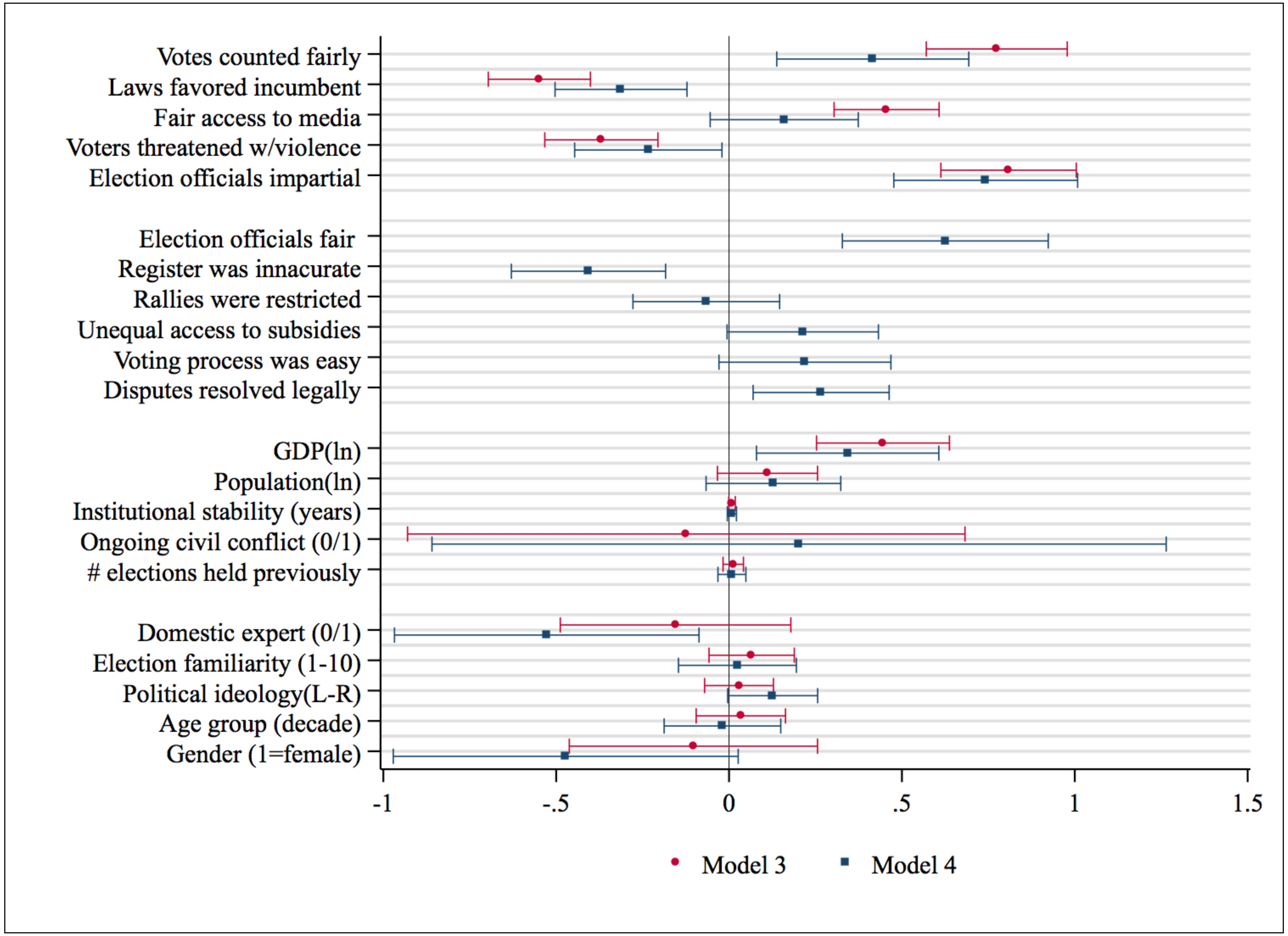
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	[illegible]
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	[illegible]
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	[illegible]
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	[illegible]
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	[illegible]
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	[illegible]

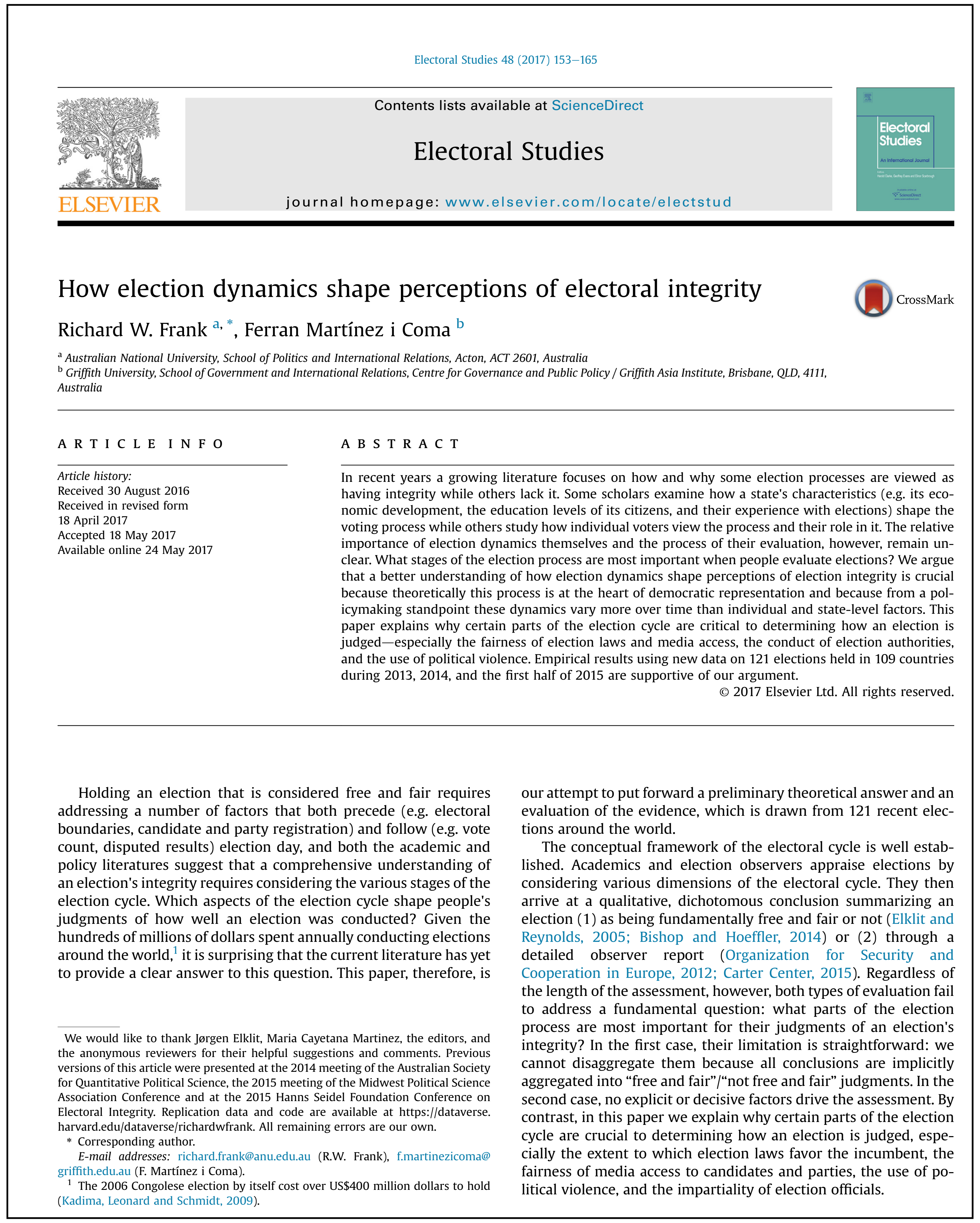
2

The **lower bound** of the confidence interval is the mean **minus** the margin of error (or two standard errors) of the mean,

The **upper bound** is the mean **plus** the mean’s margin of error.

2





Electoral Studies 48 (2017) 153–165

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

How election dynamics shape perceptions of electoral integrity

Richard W. Frank [a,*], Ferran Martínez i Coma [b]

[a] Australian National University, School of Politics and International Relations, Acton, ACT 2601, Australia
[b] Griffith University, School of Government and International Relations, Centre for Governance and Public Policy / Griffith Asia Institute, Brisbane, QLD, 4111, Australia

A R T I C L E I N F O

Article history:
Received 30 August 2016
Received in revised form
18 April 2017
Accepted 18 May 2017
Available online 24 May 2017

A B S T R A C T

In recent years a growing literature focuses on how and why some election processes are viewed as having integrity while others lack it. Some scholars examine how a state's characteristics (e.g. its economic development, the education levels of its citizens, and their experience with elections) shape the voting process while others study how individual voters view the process and their role in it. The relative importance of election dynamics themselves and the process of their evaluation, however, remain unclear. What stages of the election process are most important when people evaluate elections? We argue that a better understanding of how election dynamics shape perceptions of election integrity is crucial because theoretically this process is at the heart of democratic representation and because from a policymaking standpoint these dynamics vary more over time than individual and state-level factors. This paper explains why certain parts of the election cycle are critical to determining how an election is judged—especially the fairness of election laws and media access, the conduct of election authorities, and the use of political violence. Empirical results using new data on 121 elections held in 109 countries during 2013, 2014, and the first half of 2015 are supportive of our argument.

Holding an election that is considered free and fair requires addressing a number of factors that both precede (e.g. electoral boundaries, candidate and party registration) and follow (e.g. vote count, disputed results) election day, and both the academic and policy literatures suggest that a comprehensive understanding of an election's integrity requires considering the various stages of the election cycle. Which aspects of the election cycle shape people's judgments of how well an election was conducted? Given the hundreds of millions of dollars spent annually conducting elections around the world,[1] it is surprising that the current literature has yet to provide a clear answer to this question. This paper, therefore, is our attempt to put forward a preliminary theoretical answer and an evaluation of the evidence, which is drawn from 121 recent elections around the world.

The conceptual framework of the electoral cycle is well established. Academics and election observers appraise elections by considering various dimensions of the electoral cycle. They then arrive at a qualitative, dichotomous conclusion summarizing an election (1) as being fundamentally free and fair or not (Elklit and Reynolds, 2005; Bishop and Hoeffler, 2014) or (2) through a detailed observer report (Organization for Security and Cooperation in Europe, 2012; Carter Center, 2015). Regardless of the length of the assessment, however, both types of evaluation fail to address a fundamental question: what parts of the election process are most important for their judgments of an election's integrity? In the first case, their limitation is straightforward: we cannot disaggregate them because all conclusions are implicitly aggregated into "free and fair"/"not free and fair" judgments. In the second case, no explicit or decisive factors drive the assessment. By contrast, in this paper we explain why certain parts of the election cycle are crucial to determining how an election is judged, especially the extent to which election laws favor the incumbent, the fairness of media access to candidates and parties, the use of political violence, and the impartiality of election officials.

We would like to thank Jørgen Elklit, Maria Cayetana Martinez, the editors, and the anonymous reviewers for their helpful suggestions and comments. Previous versions of this article were presented at the 2014 meeting of the Australian Society for Quantitative Political Science, the 2015 meeting of the Midwest Political Science Association Conference and at the 2015 Hanns Seidel Foundation Conference on Electoral Integrity. Replication data and code are available at https://dataverse.harvard.edu/dataverse/richardwfrank. All remaining errors are our own.

* Corresponding author.
E-mail addresses: richard.frank@anu.edu.au (R.W. Frank), f.martinezicoma@griffith.edu.au (F. Martínez i Coma).

[1] The 2006 Congolese election by itself cost over US$400 million dollars to hold (Kadima, Leonard and Schmidt, 2009).

2

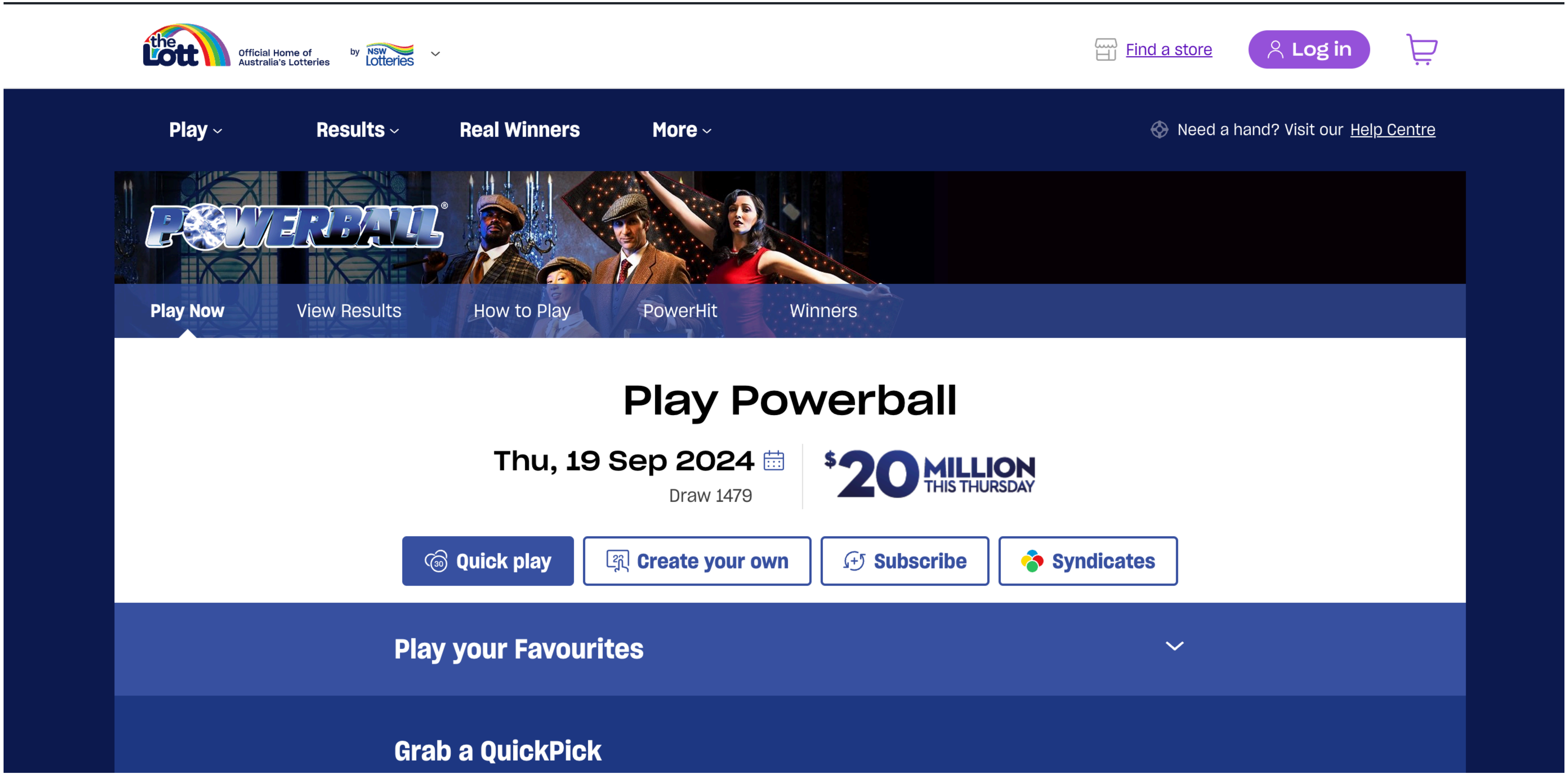
Expected value [E(P)] of the Squid Game prize (~$49 million AUD).

E(P) = (prize)*(probability of winning)

E(P) = ($49,185,319.58)*(.002 [1/456])

E(P) = $107,862

2



Official Home of Australia's Lotteries
by NSW Lotteries
Find a store
Log in
Play
Results
Real Winners
More
Need a hand? Visit our Help Centre
POWERBALL
Play Now
View Results
How to Play
PowerHit
Winners
Play Powerball
Thu, 19 Sep 2024
Draw 1479
$20 MILLION THIS THURSDAY
Quick play
Create your own
Subscribe
Syndicates
Play your Favourites
Grab a QuickPick

2

Expected value of this Thursday’s Powerball ($20 million AUD) is E(P).

E(P) = (prize)*(probability of winning)

E(P) = ($20,000,000.00)*(.00000000744 [1/134,490,400])

E(P) = $0.15

2

Probabilities involve **uncertainty**.

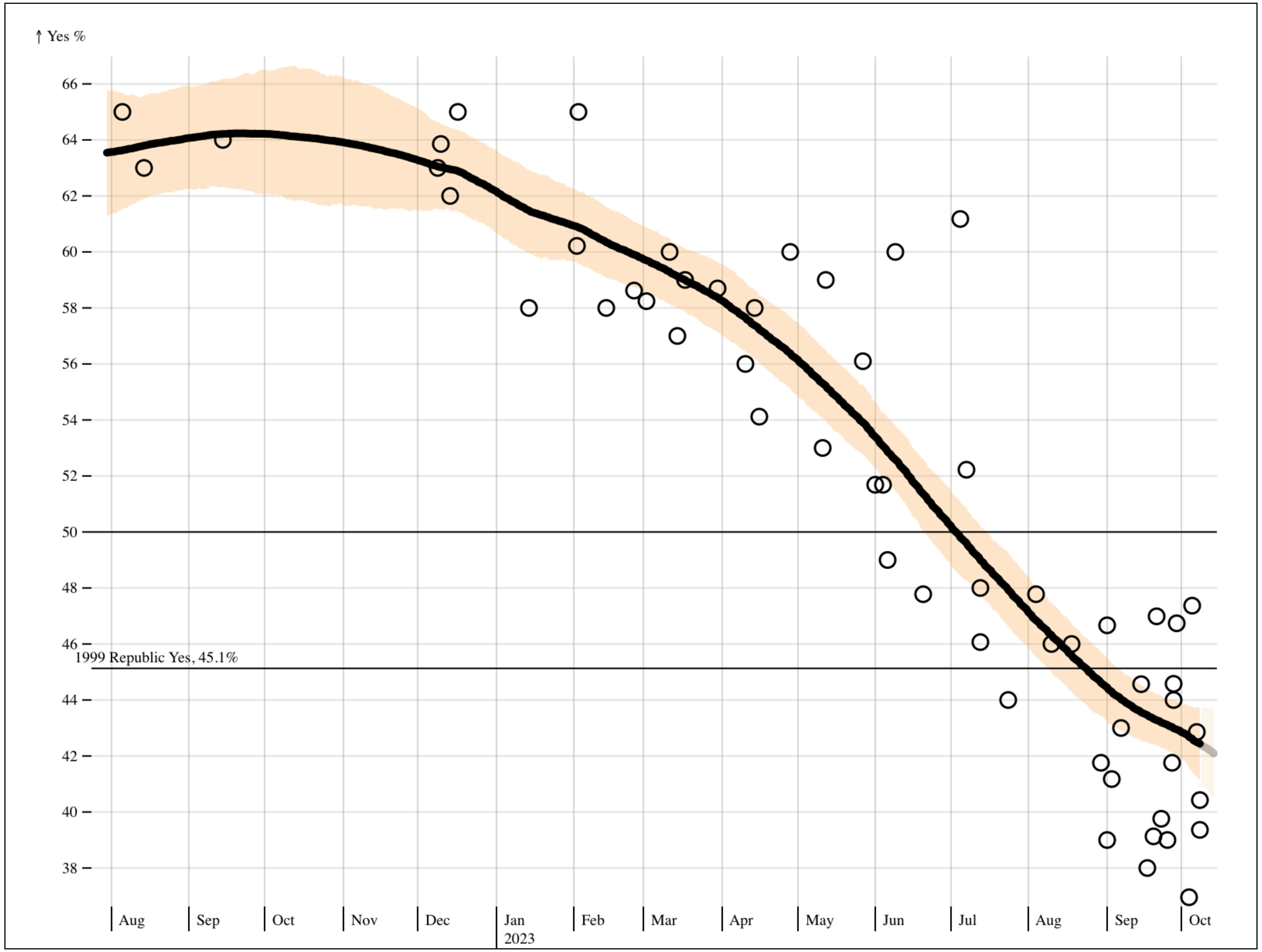
Political scientists need estimates of uncertainty as we have **sample data** instead of population data.

Probability theory comes with important **assumptions**, **strengths**, and **weaknesses**.

It will be largely relevant to us when determining **statistical significance**.

3

3



Source: Simon Jackman (https://simonjackman.github.io/poll_averaging_voice_2023/poll_averaging.html)

3

Support for Voice to Parliament (trend)

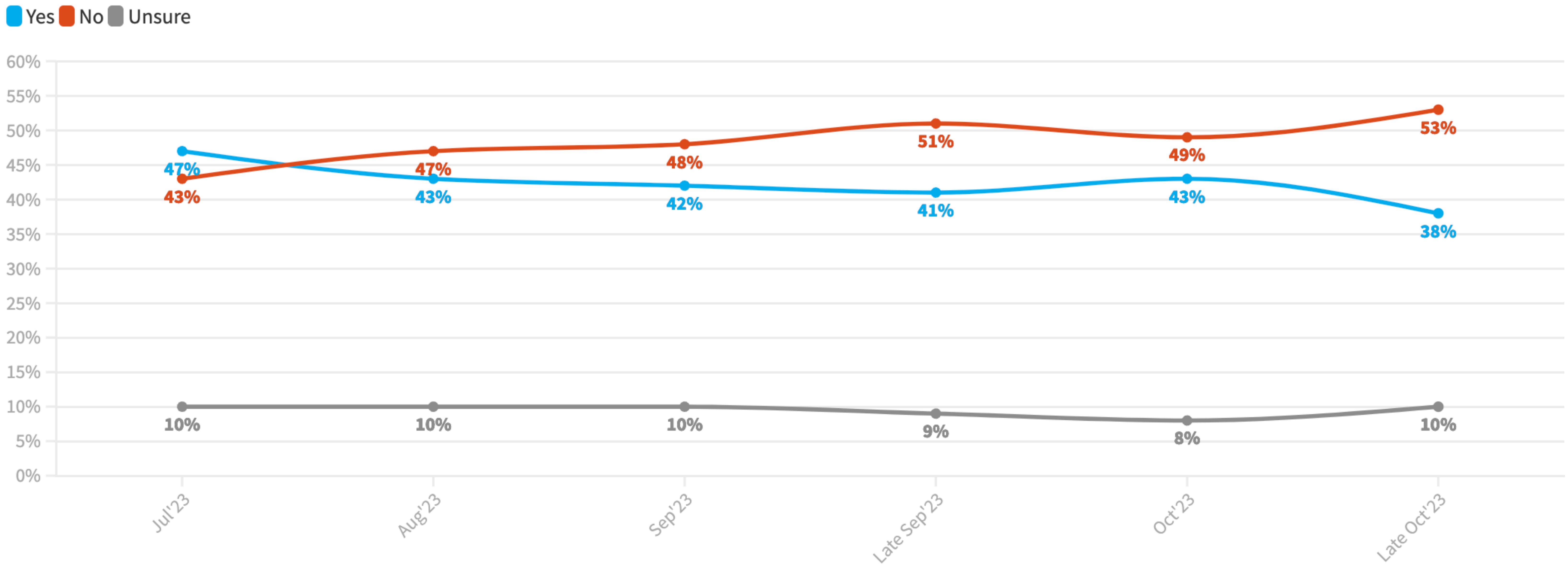
Q. (IF ALREADY VOTED) What was your vote at the referendum?

(IF NOT YET VOTED) As you may be aware, there will be a referendum held later this year on whether a Voice to Parliament for Aboriginal and Torres Strait Islander people should be enshrined in the constitution. The question at the 2023 referendum will be:

A Proposed Law: to alter the Constitution to recognise the First Peoples of Australia by establishing an Aboriginal and Torres Strait Islander Voice.

Do you approve this proposed alteration?

(IF UNSURE) Well which way are you currently leaning towards voting?

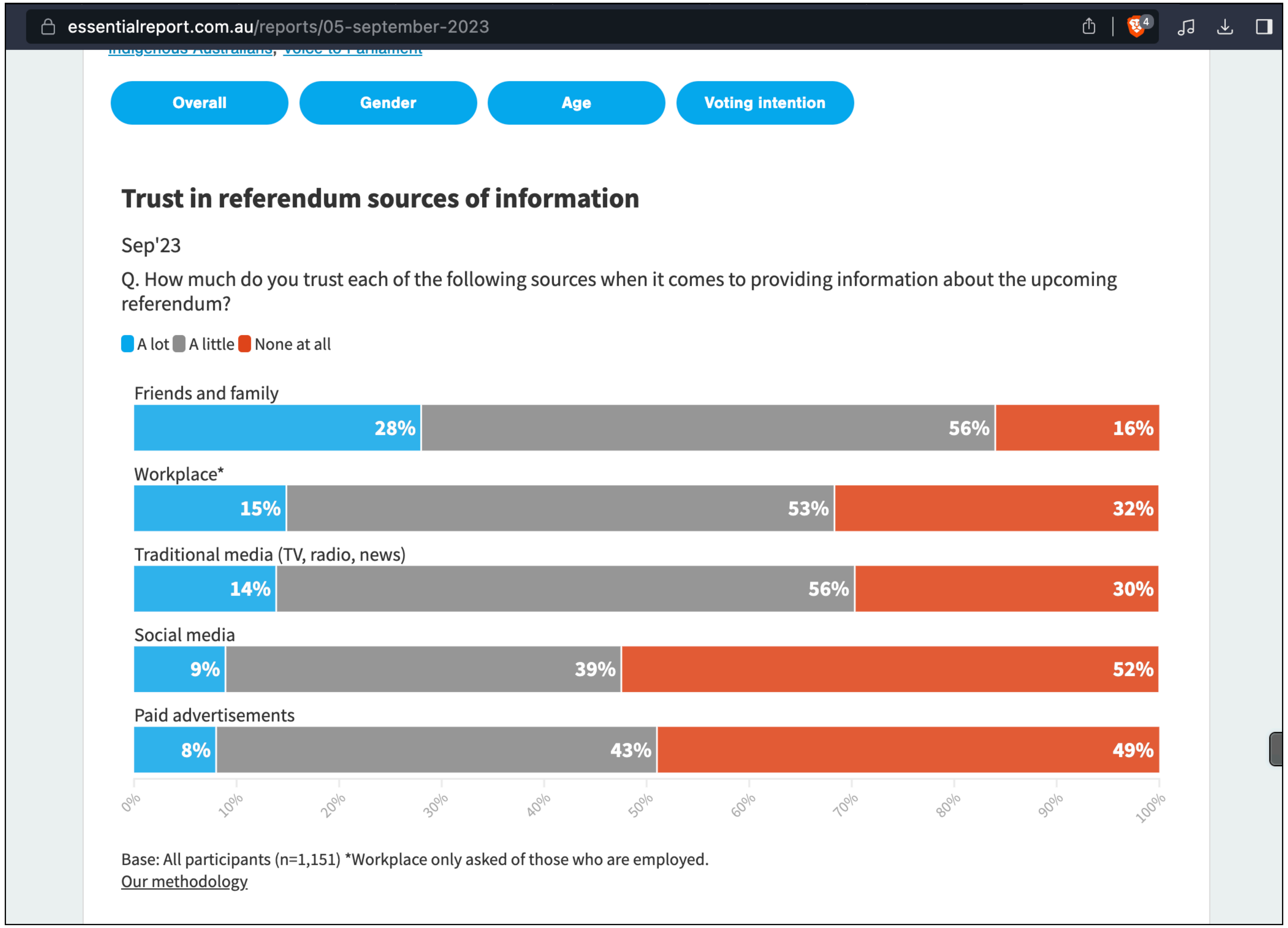


Base: All participants (min. n=1,100)

Our methodology

Source: https://essentialreport.com.au/tag/voice-to-parliament

3



Source: https://essentialreport.com.au/reports/05-september-2023

3

essentialvision.com.au/about-this-poll

Essential. RESEARCH

Home About this poll Contact Rea

About this poll

Methodology Disclosure Statement

The data gathered for this report is gathered from a fortnightly online omnibus run by Essential Research with sample provided by Qualtrics from online panels.

The online omnibus is active from the Wednesday night of each week and closed on the following Sunday. The target population is all Australian residents aged 18+. Participants were invited to participate and completed the survey online without an interviewer present and incentives are offered for participation.

The response rate varies each week, but usually delivers 1000+ interviews. Quotas are applied to be representative of the target population by age, gender and location.

RIM weighting is applied to the data using information sourced from the Australian Bureau of Statistics (ABS) and Australian Electoral Commission (AEC). The factors used in the weighting are age, gender, location and party ID.

Information for the weighting efficiency, effective sample size and margin of error for each poll (from June 2021) can be found here.

3

Individual Survey Details

Publication Date	Fieldwork Start Date	Fieldwork End Date	Target sample	Sample Size	Weighting efficiency	Effective Sample Size	Margin of Error
24-May-21	19-May-21	23-May-21	National	1,100	91%	1,006	±3.1%
07-Jun-21	02-Jun-21	06-Jun-21	National	1,104	92%	1,019	±3.1%
21-Jun-21	16-Jun-21	20-Jun-21	National	1,087	94%	1,017	±3.1%

3

How can we connect our knowledge of probability to better understand polling results?

By converting the polling results into measures of confidence that the population mean is within a certain range around the sample mean.

The **margin of error** is **half the width** of the confidence interval.

The **confidence interval** is thus twice the margin of error centred on the sample mean.

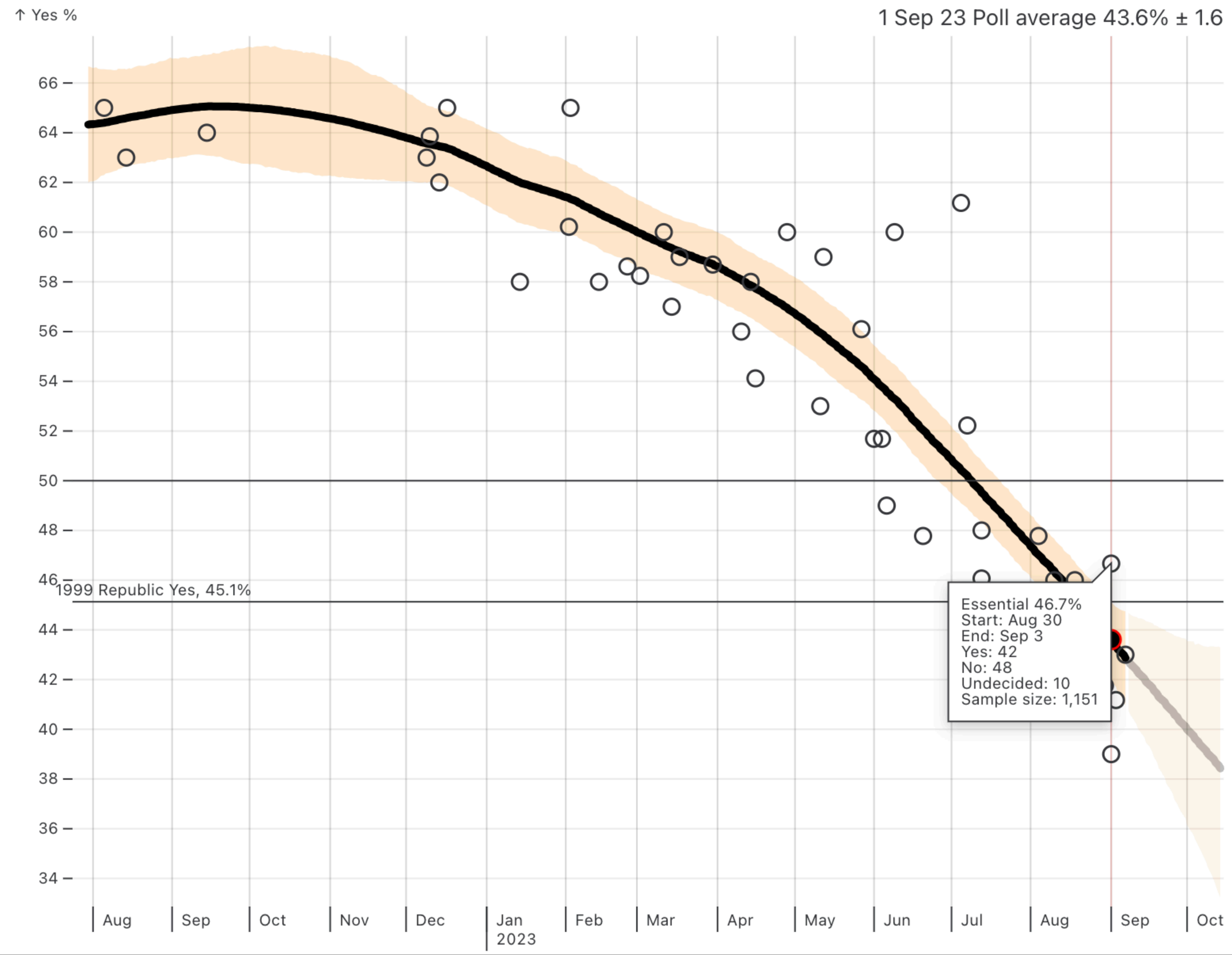
3

simonjackman.github.io/poll_averaging_voice_2023/poll_averaging.html

magnitude of the trend, even under the assumptions encoded in the model.

With these qualifications and caveats in mind, the model forecast of the poll average on October 14 is 38.4% ± 5.0.

Also included on the graph are two horizontal reference lines, one at 50% and another at 45.1%, the national “Yes” vote in the Republic referendum in 1999.



So if the Sept 2023 poll suggests that 48% no, 42% yes, 10% undecided with a 3.1% margin of error…

Response	Percentage	Lower 95% bound	Upper 95%
No	48	44.9	51.1
Yes	42	38.9	45.1
Undecided	10	6.9	13.1

3

JOURNAL OF AUSTRALIAN STUDIES
2023, VOL. 47, NO. 2, 373–397
https://doi.org/10.1080/14443058.2023.2175892

Routledge
Taylor & Francis Group

Check for updates

Support in the Polls for an Indigenous Constitutional Voice: How Broad, How Strong, How Vulnerable?

Murray Goot

Politics and International Relations, Macquarie University, Sydney, Australia

ABSTRACT
Following the prime minister's announcement, in May 2022, that Australians would be asked to decide whether to have an Indigenous Voice to Parliament inscribed in the Constitution, a large number of polls sought to measure the breadth and strength of support for a constitutionally enshrined Voice. Some also sought to measure the appeals that might make support for a Voice either more attractive or more vulnerable. This article shows that support for a constitutional amendment, while broad, was not strong: that while majorities were in favour of change—nationally and in most states—there was no majority strongly committed to change, and the majority in favour of constitutional change was declining. It shows that while most Labor voters and the Greens supported the change, Coalition supporters increasingly did not. And it shows which considerations appeared to resonate with respondents and which did not. In the course of documenting and analysing these findings, this article offers a critique of the polls: the wording and sequencing of some of the questions, some of the response options, and the questions not asked.

KEYWORDS
Indigenous recognition; constitutional referendums; Australian politics; opinion polls; Voice to Parliament

The election of a Labor Government in May 2022, committed to holding a referendum on an Indigenous Voice in the Constitution during its first (and possibly only) term, was followed by a flurry of public opinion polls in the first stage of what was destined, effectively, to be one of the longest referendum campaigns in Australia's history. By early 2023, the number of questions on an Indigenous issue that had made their way into the press almost certainly exceeded the number asked in any year since 2000, the end of the Decade of Reconciliation.[1] That almost all the polling on a Voice assumed a referendum would precede and authorise legislation, and that a Voice would not simply be legislated, speaks to the way political agendas—and polling agendas—are shaped. Polls are important: as interpreted by the press, by the parties and by other political actors, they affect public expectations, partisan manoeuvring, and political debate.

No fewer than 10 pollsters—more than have tried their hand at gauging party support ahead of any Australian election—attempted to ascertain support for a constitutionally inscribed Voice. Some tried to gauge how strongly respondents were committed to

CONTACT Murray Goot murray.goot@mq.edu.au Politics and International Relations, Macquarie University, Balaclava Road, Macquarie Park, NSW 2109, Australia

[1]Murray Goot and Tim Rowse, *Divided Nation? Indigenous Affairs and the Imagined Public* (Carlton, VIC: Melbourne University Press, 2007), 19–20.

3

AEC Australian Electoral Commission

TALLY ROOM

Skip to main content

2022 Federal Election

House of Representatives +

Results by division

Candidate search

Further information & results

Downloads

Senate +

Home / House of representatives / South Australia / Adelaide / Adelaide East

Updated: Mon, 13 Jun 2022 4:23:43 PM AEST

Adelaide East - polling place

- Christian Brothers College (Senior Campus), 214 Wakefield St, ADELAIDE SA 5000

?

First preferences

Candidate	Party	Votes	%	Swing (%)
GRANTHAM, Amy	Liberal	311	29.53	-8.36
GEORGANAS, Steve	Australian Labor Party	359	34.09	-1.35
ALLWOOD, Sean	United Australia Party	26	2.47	+1.62
McMILLAN, Matthew	FUSION: Science, Pirate, Secular, Climate Emergency	21	1.99	+1.99
GALDIES, Rebecca	The Greens	302	28.68	+7.96
GERHARD, Faith	Australian Federation Party	6	0.57	+0.57
ALLWOOD, Gayle	Pauline Hanson's One Nation	28	2.66	+2.66
......	Australian Democrats	0	0.00	-1.78
......	Animal Justice Party	0	0.00	-3.31
Formal		**1,053**	**97.05**	**-1.11**
Informal		**32**	**2.95**	**+1.11**
Total		**1,085**	**0.83**	**-0.16**

3

Advance Access publication March 12, 2012

Political Analysis (2012) 20:211–234
doi:10.1093/pan/mps003

What the Numbers Say: A Digit-Based Test for Election Fraud

Bernd Beber and Alexandra Scacco
Department of Politics, New York University, 19 West 4th Street, New York, NY 10012
e-mail: bernd.beber@nyu.edu (corresponding author), alex.scacco@nyu.edu

Edited by R. Michael Alvarez

Is it possible to detect manipulation by looking only at electoral returns? Drawing on work in psychology, we exploit individuals' biases in generating numbers to highlight suspicious digit patterns in reported vote counts. First, we show that fair election procedures produce returns where last digits occur with equal frequency, but laboratory experiments indicate that individuals tend to favor some numerals over others, even when subjects have incentives to properly randomize. Second, individuals underestimate the likelihood of digit repetition in sequences of random integers, so we should observe relatively few instances of repeated numbers in manipulated vote tallies. Third, laboratory experiments demonstrate a preference for pairs of adjacent digits, which suggests that such pairs should be abundant on fraudulent return sheets. Fourth, subjects avoid pairs of distant numerals, so those should appear with lower frequency on tainted returns. We test for deviations in digit patterns using data from Sweden's 2002 parliamentary elections, Senegal's 2000 and 2007 presidential elections, and previously unavailable results from Nigeria's 2003 presidential election. In line with observers' expectations, we find substantial evidence that manipulation occurred in Nigeria as well as in Senegal in 2007.

1 Introduction

Suppose you have been asked to assess how ''clean'' a past national election was in different areas of a country. You have poor national-level information about the make-up of the voting population and virtually no information at the subnational level. You do not have access to results from any previous elections. Constituency maps are either not publicly available or do not exist. In essence, the only information you have is a list of electoral returns. This is a situation election monitors are likely to encounter in a range of developing countries where fraud may occur, and it is a situation in which regression-based tests for outliers will not work. Is it possible to say, with some confidence, whether results have been manipulated by looking at the return sheets only?

We approach this problem by developing a digit-based test that exploits human biases in number generation.[1] We refer to four expectations from the relevant psychology literature: First, fair election procedures should produce returns where last digits occur with equal frequency, but laboratory experiments have shown that individuals tend to disproportionately select particular numerals, even when they have incentives to properly randomize. Second, individuals tend to underestimate the likelihood of digit repetition in sequences of random integers, which means that we should observe relatively fewer instances of repeated numbers on manipulated vote report sheets. Third, laboratory experiments demonstrate a preference for pairs of adjacent digits, which suggests that such pairs should be abundant on fraudulent return sheets. Fourth, subjects avoid pairs of distant numerals (i.e., digits that are neither repetitive nor adjacent), so those should appear with lower frequency on manipulated returns.

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.
[1]In the course of the review process, we became aware of a set of papers that independently developed a related test in the field of research ethics (Mosimann, Wiseman, and Edelman 1995; Mosimann and Ratnaparkhi 1996; Mosimann et al. 2002). We provide alternative theoretical foundations, a set of tests for trailing digit pairs as opposed to terminal digits alone and novel data. Additional information is available in the supplementary materials posted on the *Political Analysis* Web site.

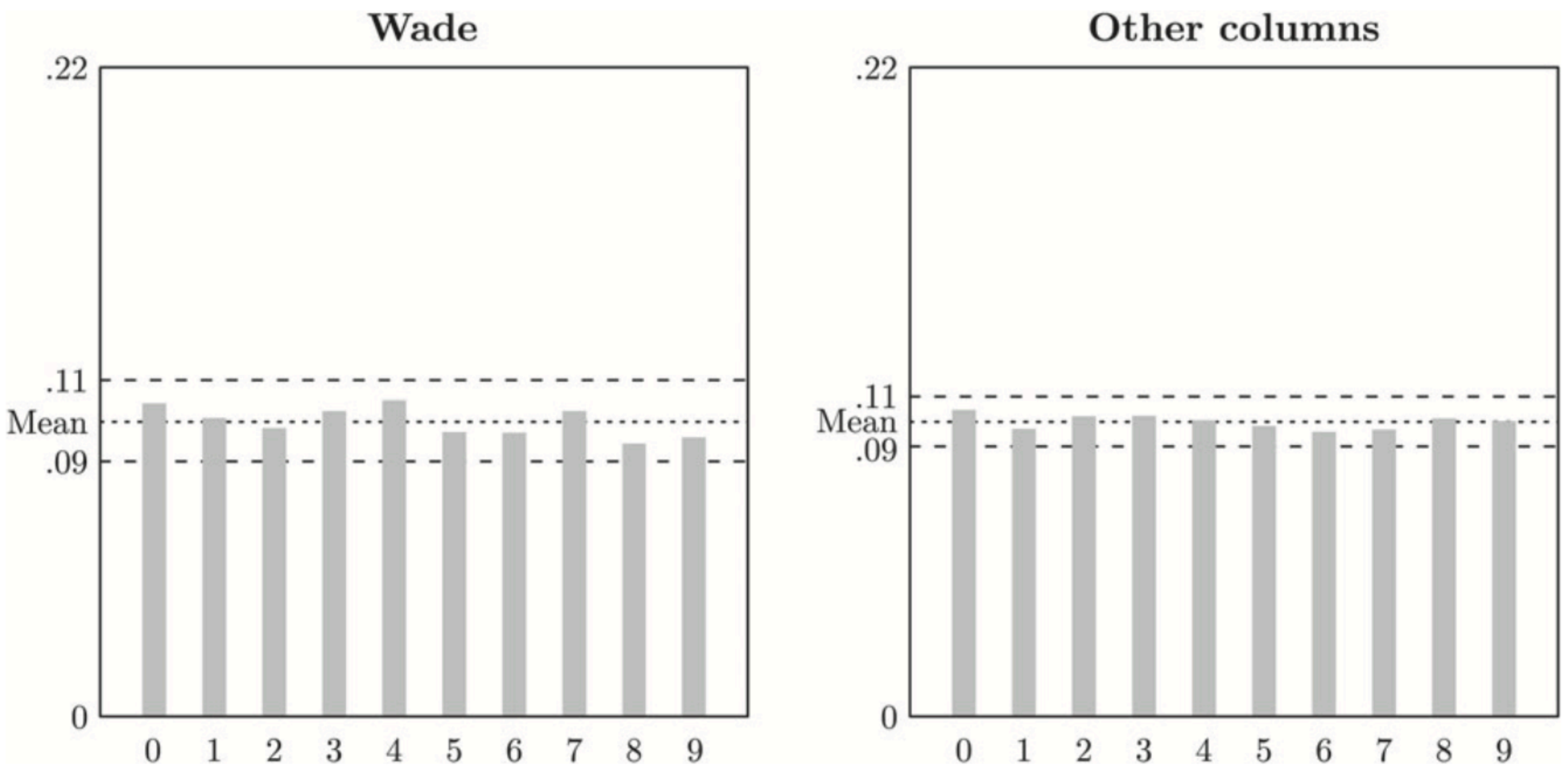


Fig. 9 Frequencies of last digits, Senegal 2000.

3

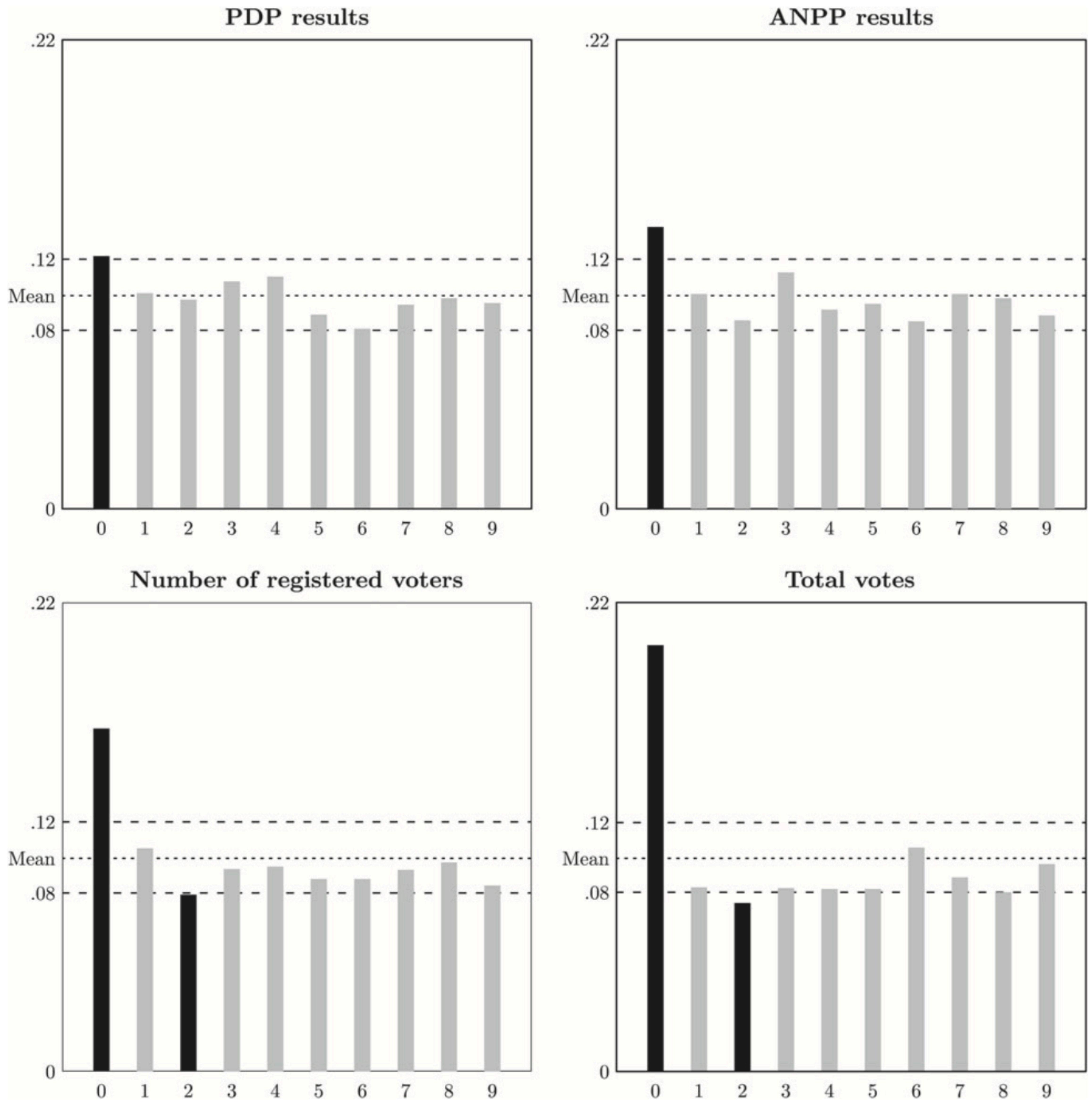


Fig. 4 Frequencies of last digits, Nigeria 2003.

3

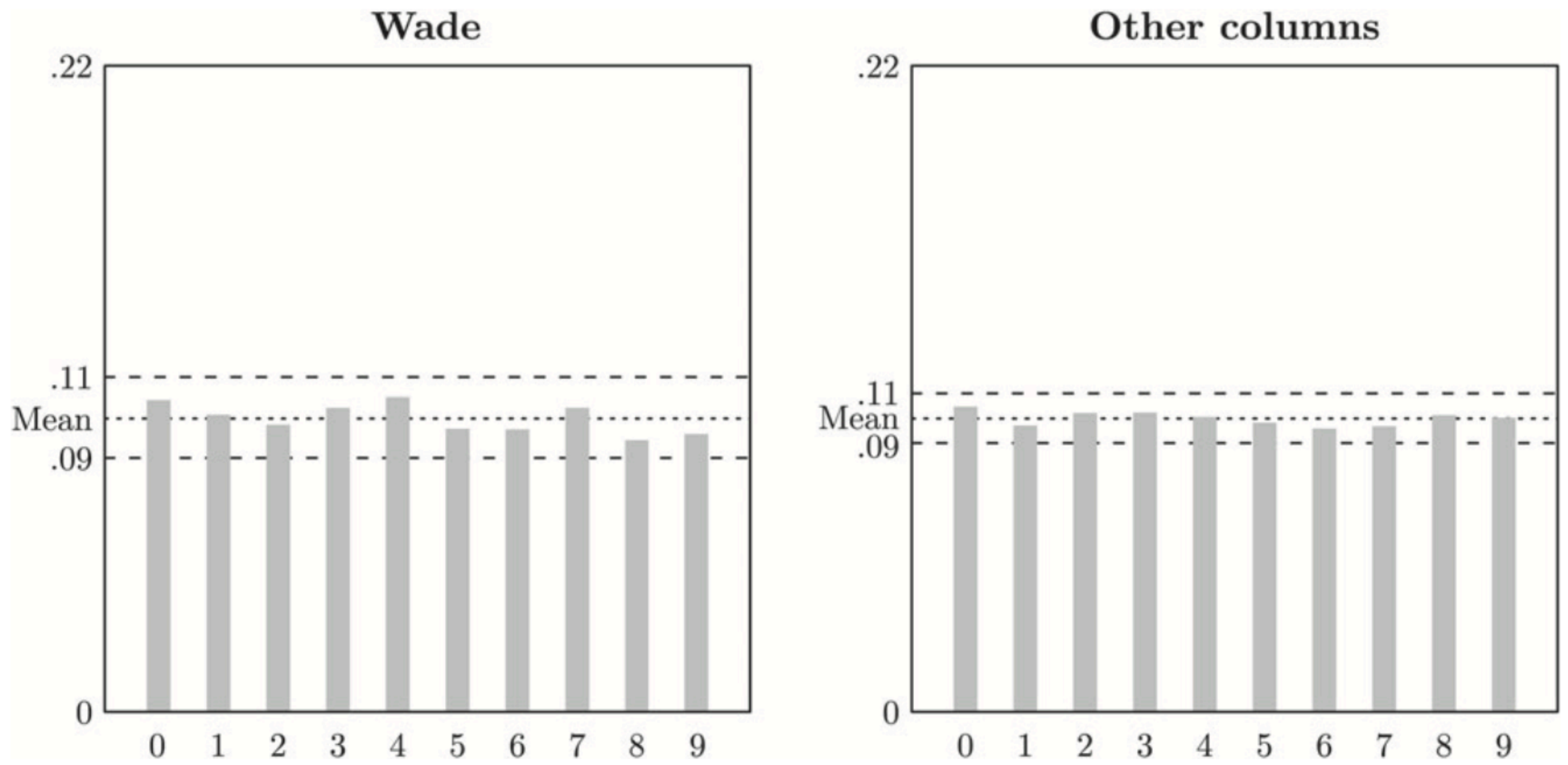


Fig. 9 Frequencies of last digits, Senegal 2000.

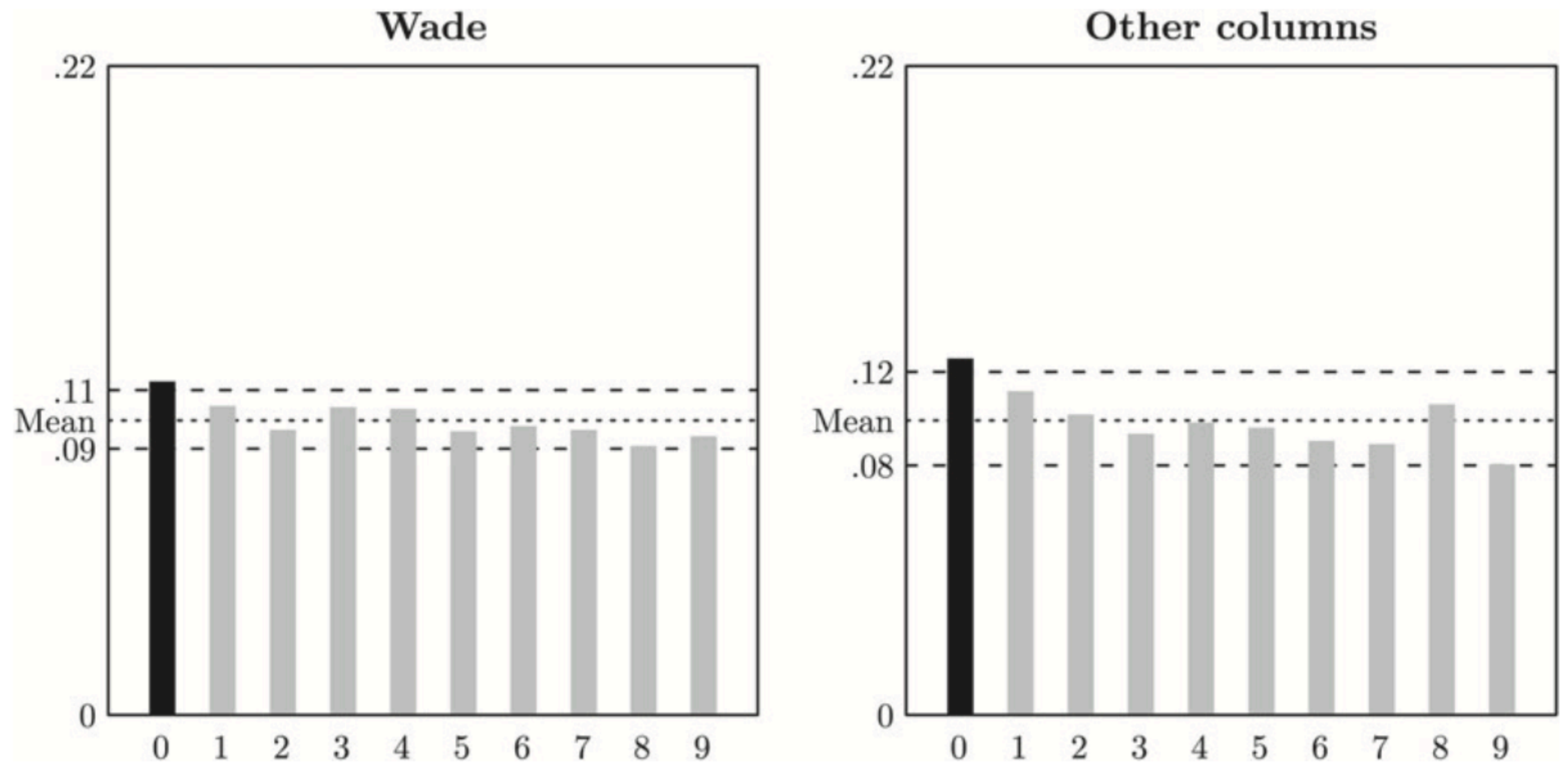
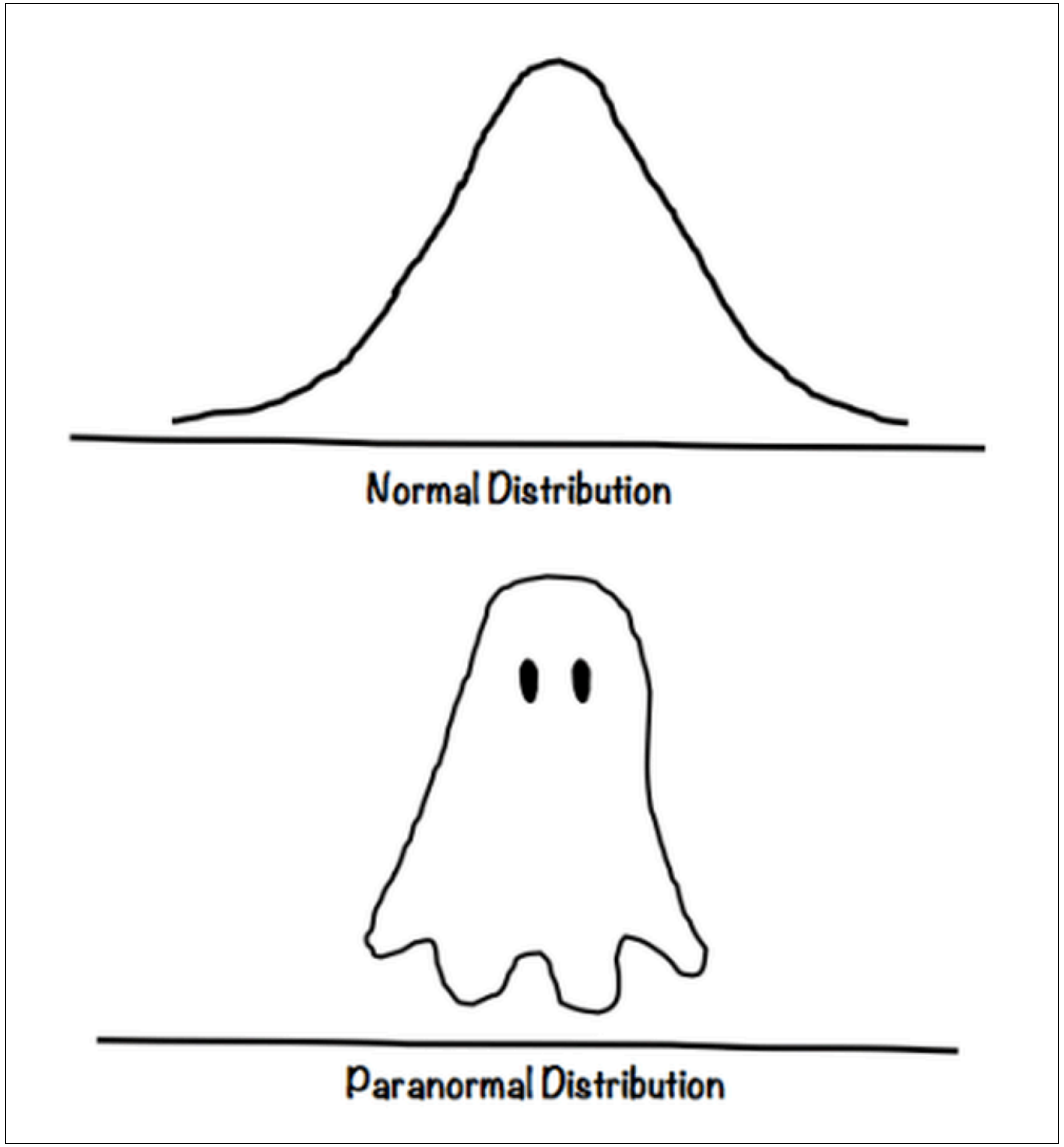


Fig. 10 Frequencies of last digits, Senegal 2007.

3



https://www.siliconrepublic.com/careers/10-statistician-memes-to-celebrate-data-science-week