

1

State and Statistics

[In Nigeria,] the census figures became strong political weapons rather than statistical data to be used for planning for socio-economic development.

—Adepoju (1981: 35)

In a severely divided society ... an election can become an ethnic head count ... a census needs to be 'won'. So the election is a census, and the census is an election.

—Horowitz (2000: 196)

Introduction

Over the past two centuries, the deep and multifaceted relation between statistics and statecraft has emerged as one of the defining features of states across the world. Modern states depend on statistics for the planning and evaluation of interventions. The growing size and complexity of operations undertaken by states have deepened their dependence upon statistics.¹ Bureaucratisation² and technocratisation³ of policymaking as well as the growing capacity of non-state actors to challenge government policies⁴ have also pushed states towards statistics.

The relationship between state and statistics is not merely instrumental though. Given their intimate relation with the origin and evolution of modern states, statistics are integral to the self-imagination of states and, also, to how they are imagined by people.⁵ In its earlier eighteenth-century sense, statistics was 'a set of administrative routines needed to describe a state and its population' (Desrosières 1998: 16), a description of the state by and for itself (ibid.: 147).⁶ By the early nineteenth century, almost all Western countries had established statistical offices (Tooze 2003: 2; Urla 1993: 821) as 'national statistics' had come to be seen 'as one of the vital attributes of the nation-states then under construction or seeking to assert themselves' (Desrosières 2013: 10).⁷ The quality of statistics produced by a country began to be seen as an attribute of its socio-economic and political

4 Numbers in India's Periphery

development, with advanced economies and liberal democracies being associated with better statistical systems (Porter 1995: 80; see also Urla 1993: 821). Around the same time, statistics also began to be seen as enablers of public interest.⁸ And the census 'became less concerned with what the people could be obliged to do for the state and more concerned with what the state could do for them' (Coleman 2012: 335) amidst an emerging 'shift towards willing participation [in state-sponsored data collection efforts] on the part of the respondents' (Bookman 2013: 51; see also Starr 1987: 12; Prewitt 2010: 239).

Democracy was another site for the intertwining of state and statistics. The origins and spread of modern democracy and state-sponsored human population censuses are closely related. The United States of America (USA), where decennial delimitation intertwined democracy and census in the late eighteenth century, was an early exemplar of this relationship. By the early twentieth century, conflicts over the delimitation of administrative units and electoral constituencies and, by implication, population censuses had almost become 'rites of passage in the lives of modern states outside the West' (Kumar 2019: 1).⁹

The long-standing and multifaceted relationship between state and statistics notwithstanding, government statistics of most (developing) countries are often not free of errors.¹⁰ Errors result from definitional and measurement problems,¹¹ as well as from capacity constraints,¹² corruption¹³ and unrealistic targets¹⁴ (Figure 1.1). Non-governmental actors in most developing countries have limited expertise to question government data, let alone build alternative databases. Restrictions

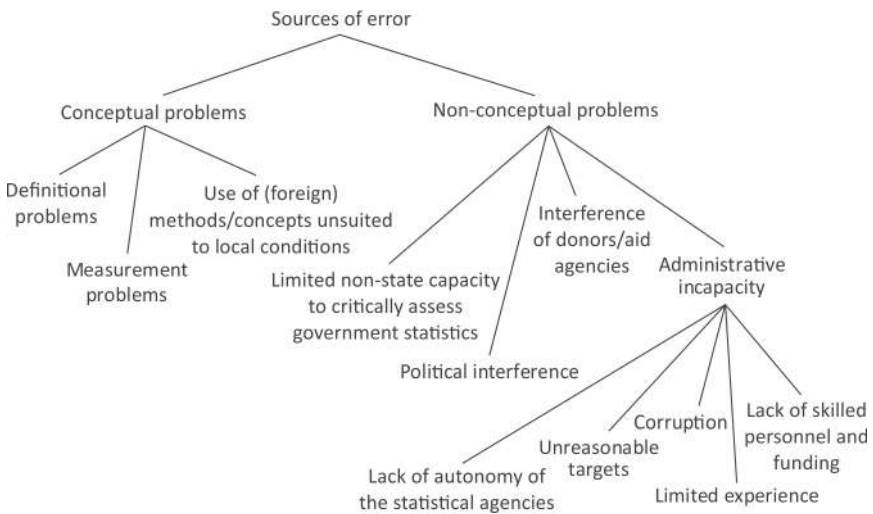


Figure 1.1 Sources of error in statistics of developing countries

Source: Authors.

on freedom of expression and media, weak judiciaries and lower rates of literacy further impede the scrutiny of government statistics in these countries. Developed countries face fewer problems because of a longer history of engagement with modern statistics, greater availability of skilled personnel and resources to build reliable databases, better quality of the general administration and greater non-governmental capacity to critically assess government statistics and build alternative databases.

Statistics are also susceptible to errors because of interference and contestation by interest groups competing for resources and power within and across countries. Countries manipulate economic statistics to attract international finance and secure preferential access to development funding. In the late 1990s, the World Bank ‘lowered’ the per capita income of China ‘in line with the Chinese request’ to allow it to access ‘very cheap, long-term loans [offered by the International Development Association, the soft-loan arm of the World Bank] available to countries whose average income is below a certain level’ (Wade 2012: 17). A decade later, the Chinese leader Le Keqiang noted that his country’s national accounts were unreliable (Ninan 2018).¹⁵ In 1987, Myanmar leader Ne Win complained that ‘political manipulation’ and ‘intentional falsification’ had rendered economic statistics useless (Steinberg 2015: 10). In Pakistan, the ‘basis for undermining the integrity of official statistics was laid in the 1990s when the Ministry of Finance engaged in systematic manipulation and misreporting of the public finances data ... to “cheat” the IMF’ (Sherani 2016). Michalski and Stoltz (2013: 591), who examine strategic falsification of economic statistics in a cross section of countries, point out that

Argentina by misstating inflation figures avoided paying out higher interest on government bonds indexed to inflation ... and raising the wages in the public sector. Greece enjoyed lower borrowing rates (close to Germany’s) on its government debt ... because investors did not know the entire extent of Greek budget troubles. (see also Coyle 2014: 3–4).

There is also a non-instrumental reason why developing countries manipulate data. These countries compete in both the real and symbolic realms. They are obsessed with global rankings as markers of status and, consequently, need to measure their progress relative to other countries, especially in the West.¹⁶ Statistics allow easy, even if context-independent and, therefore, less meaningful, comparisons. ‘India and China are’, for instance, ‘among the 60-plus countries that have government units dedicated to moving upwards [on the World Bank’s ease of doing business ranking], almost as if it were an end in itself’ (*Economist* 2018). Myanmar’s military regime manipulated economic statistics ‘to prove that their country was every bit as tigerish as its Asian rivals’ (Cockett 2015: 158).

6 *Numbers in India's Periphery*

The Developing World's Experience

Most liberal democracies in the West have not seen deeply divisive ethnopolitical contests over census after the Second World War.¹⁷ However, census has proven to be an arena of conflict across Asia and Africa as communities/regions indulge in competitive manipulation of statistics to secure favourable delimitation of administrative units and electoral constituencies and, by implication, a larger share of the public pie. The political history of independent Nigeria aptly illustrates the experience of developing countries. Adepöju (1981: 29, 35) points out that

population issues play dominant roles and are largely responsible for major landmarks in the contemporary Nigerian political scene: it precipitated the constitutional crisis in the country in 1962; it played a major role in the crisis in the old Western Nigeria in 1965; was largely responsible for the military take-over in 1966; contributed greatly to the fall of Gowon's regime in 1975 and still looms large in the minds of Nigerians with the recent demand for the creation of even more states in the country; and revenue allocation among existing states soon after the return to civilian rule ... [ultimately,] the census figures became strong political weapons rather than statistical data to be used for planning for socio-economic development. (For a related discussion on Ethiopia, see *Economist* 2019)

In several countries, governments collect limited or no information on ethnicity, religion and language to avoid conflict (Horowitz 2000; Kertzer and Arel 2002b; Bookman 2013).¹⁸ In others, governments delay the release of results.¹⁹ In still other cases, data collection exercises are either indefinitely suspended to eliminate a source of conflict or cannot be conducted because rival groups resist the authority of the government to conduct surveys in their strongholds. Lebanon presents an extreme example in this regard. It did not conduct a census after 1932 as it feared that 'taking one would reveal such changes in the religious composition of the population as to make the marvelously intricate political arrangements designed to balance sectarian interests unviable' (Geertz 1973: 275; see also Horowitz 2000: 195).

Most countries in India's neighbourhood have not been able to conduct national censuses regularly.²⁰ The decennial cycle of census was disrupted in Pakistan when it postponed the 1971 Census. Pakistan has not been able to conduct censuses regularly since then. It postponed the census five times between 1991 and 1998 fearing conflict in case the demographic balance among ethnolinguistic groups was altered and ultimately managed to produce figures that affirmed the status quo (Weiss 1999: 687, 691; see also Khan 1998: 481). In 2012, Sri Lanka conducted its first countrywide census in three decades (Government of Sri Lanka 2012). Myanmar conducted a census in 2014 after a gap of 31 years, but once again, it did not adequately cover the minorities (TNI-BCN 2014; Aung 2018). Afghanistan

could not complete the 1979 Census due to political turmoil and has not been able to conduct a national census since then (Tolo News 2018). China did not conduct a census between 1964 and 1982 (*Global Times* 2011).

Africa abounds in examples of abandoned and delayed censuses (Lalasz 2006).²¹ Even major African countries such as Nigeria have not been able to conduct regular and reliable censuses (Okolo 1999). Nigeria's 1962 Census had to be repeated in 1963 due to concerns about the quality of data. The next census was conducted in 1973, but it too proved to be controversial and the results were not published. Nigeria delayed censuses for a long time, fearing unrest and eventually conducted censuses that almost reproduced earlier federal population shares (Fawehinmi 2018).

In a survey of ethnic conflicts across the developing world, Horowitz highlights the mutually constitutive relation among ethnic conflict, elections and competitive manipulation of census:

As an entitlement issue, the census is a splendid example of the blending of group anxiety with political domination.... Disputes over census results in ethnically divided societies are common.... In a severely divided society ... an election can become an ethnic head count. Now it is clear that a census needs to be 'won.' So the election is a census, and the census is an election. (Horowitz 2000: 194–6)²²

Political considerations also affect sample surveys. In late colonial Ghana, the choice of the geographic scope of household budget surveys 'closely mirrored the political interests of those in power' (Serra 2014: 10).²³ Deaton and Kozel (2005: 190, 196) hint at the politically contested nature of India's National Sample Surveys (NSSs):

The history of poverty lines in India is a case study in the interaction of science and politics, with political decisions often claiming a scientific basis.... There is no suggestion here that the statistical failures in India in the 1990s were the result of undue interference by politicians or policymakers in data collection or publication. Yet politics in the broad sense played a role. In evaluating the reforms, the political right had an interest in showing low poverty, and the political left in showing high poverty, and this undoubtedly intensified the debate on survey design and led to the unfortunate compromise design that temporarily undermined the poverty monitoring system.²⁴

The contestations of the 1990s played out again towards the end of the following decade. The 66th round of the NSS (2009–10) proved to be controversial because it showed that employment generation fell significantly short of the target of the Eleventh Five Year Plan. This contradicted the presumptions of '[s]ome highly placed officials,' who instead of 'questioning their own priors ... decided that

8 Numbers in India's Periphery

the data must be wrong, and castigated the NSSO for its faulty investigative methods' (Chandrasekhar and Ghosh 2011; see also *EPW* 2011). The then deputy chairman of the Planning Commission repeatedly argued that the results of the latest sample survey 'were not the best judge of the extent of impact government policies have had on the poor because it was a drought year' (*Economic Times* 2012; 2013a).²⁵ The 66th round was followed by another quinquennial survey in 2011–12. Chandrasekhar and Ghosh (2013) argue that

for the UPA [United Progressive Alliance government], the 2009–10 figures must have been particularly disappointing.... It is possibly for this reason, and the fact that waning growth required finding other indicators to place before the electorate in 2014,²⁶ that the NSSO was encouraged to break from tradition and generate one more, large sample survey of employment relating to 2011–12.²⁷

We have so far discussed select instances of politicisation of statistics from across the world. Different local factors explain the intertwining of politics and statistics in different countries. However, common to all of them 'is a constitutive interrelationship between quantification and democratic government' because numbers are 'intrinsic to the forms of justification that give legitimacy to political power in democracies' and 'integral to the technologies that seek to give effect to democracy as a particular set of mechanisms of rule' (Rose 1991: 675). In such settings, political and numerical controversies become inseparably intertwined and political disputes are often waged in the language of numbers (*ibid.*: 685). For instance, 'Arguments about numerical quotas, availability pools and demographic imbalance become a substitute for democratic discussion of the principles of equity and justice' (Kenneth Prewitt quoted in *ibid.*: 680). Since the process of creating equal opportunity conditions for minority/underprivileged groups is slow and uncertain, 'statistical proportionality' becomes 'a favored legal and administrative tool' (Prewitt 2003: 16). While our focus is on democratic states, the discussion can be extended to other kinds of states too.²⁸

The conceptual vocabulary, categories and methods of measurement and modes of dissemination of data represent the government's political priorities and the balance of power in the society (notes 18 and 23 of this chapter). The legal–institutional architecture of government statistical infrastructure is negotiated 'between stakeholders from ... politics, the economy and law' (Heine and Oltmanns 2016: 205). Political elites are, however, not necessarily committed to the total welfare of the society and

are likely to perceive four specific roles and objectives for official statistics: 1) a policymaking tool; 2) a signal to supporters and detractors; 3) a means for enhancing control over rents and revenues; and 4) a source of pressure and

competition to their position.... In addition, two factors influence elite decision-making on official statistics: their access to sources of financing and their political ideology. (Krätke and Byiers 2014: 26; see also Heine and Oltmanns 2016: 203, 207; Taylor 2016: 12–13).²⁹

In short, statistics are sites of political contestation insofar as *what to count, how to count* and *how to use statistics* are inherently political choices. Or, as Bookman (2013: 71) puts it, ‘who counts depends on who counts’. While it is true that statistics often serve elite interests, ‘[i]n the hands of the socially or politically disenfranchised, numbers may also be a language of social contestation, a way that ethnic groups, women, and minorities can make themselves visible, articulate their “differences” from the dominant society, and make claims upon the state and its services’ (Urla 1993: 818; see also Goderle 2016: 87; Taylor 2016: 13). ‘[A]t stake in minority concerns with statistics are not only competing claims to resources but also competing claims to truth’ because ‘[a]s part of a modern regime of truth that equates knowledge with measurement, statistics occupy a place of authority in contemporary modes of social description; they are technologies of truth production’ (Urla 1993: 819).

Scope and Framework

Statistics of interest to social sciences are inherently prone to errors because of the difficulty in conducting designed experiments and the intrinsic randomness of human behaviour. We will, however, focus on erroneous statistics that are a product of deliberate human choices. We will examine errors engendered by deliberate intervention in the generative processes of government statistics and not merely their misuse or misinterpretation, which happens *after* statistics have already been collected. We will, however, not deal with conflicts over the choice of modes of statistical reasoning that can affect inferences derived from data (Schweber 2001) or with politically motivated changes in government policies that can alter the data generation process (Gregg 1994). Furthermore, our discussion is restricted not only to intra-national statistics but also to statistics produced and published domestically.³⁰ Even within the latter we do not deal with administrative statistics, that is, statistics generated by government departments as by-products of their routine operations or through statutory administrative returns (GoI n.d.16: para 14.3.1).³¹ Chapter 4, though, discusses the quality of select administrative data that are compared with census data.

We will examine systematic and deliberate errors in key government statistics such as area, population and monthly per capita consumer expenditure (MPCE).³² Area statistics and maps are essential for conducting population censuses.³³ The territory covered, along with any changes in its area in successive censuses, should

10 Numbers in India's Periphery

be clearly and explicitly stated' because 'population figures have no meaning unless they refer to a well defined territory' (GoI 2009b: 5). The first few questions/entries in any census schedule help to uniquely and geographically locate the respondent. In fact, the primary, even if inadequately acknowledged, role of Houselisting and Housing Census – the first round of Indian censuses – is to help partition and map the territory into enumeration blocks consisting of numbered buildings.³⁴ Once data are collected, maps (and area statistics) are needed again to delimit and display them.

Maps and population are in turn essential for collecting most other statistics. Area and population are needed to calculate per unit estimates. Furthermore, population census provides information for planning sample surveys, identifying the sampling frame and estimating sample statistics (Kish and Verma 1986) and also provides 'benchmark data for evaluating the reasonableness of the over-all survey results' (GoI 2009b: 12). Maps are needed to locate sub-samples on the ground.

Most discussions on data quality examine statistics in isolation, overlooking the interrelationships between different kinds of statistics as well as the larger context of data deficit. The context can be understood if we locate *data* deficit, which includes both the unavailability and poor quality of data, in relation to *democracy* and *development* deficits. Figure 1.2 identifies some of the key interlinkages among the three.

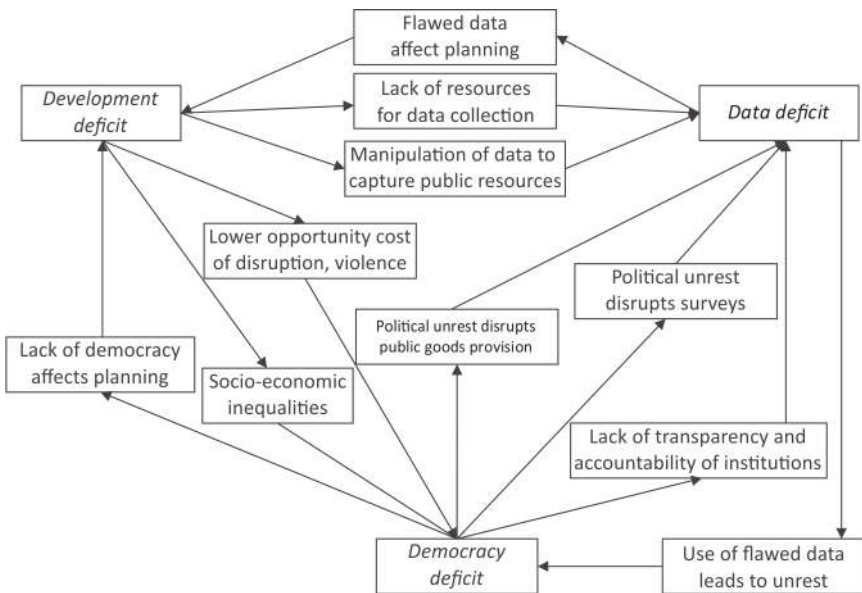


Figure 1.2 A triple deficit
 Source: Authors.

It is well known that political unrest/instability is both a cause and consequence of underdevelopment (Alesina et al. 1996; Gyimah-Brempong and Traynor 1999), that is, *democracy*³⁵ and *development* deficits are interrelated. Furthermore, the availability and quality of data are affected by and affect democratic processes. The absence or weakness of democracy is associated with less transparent and less accountable institutions. Political conflict engendered by the breakdown of democratic institutions that can credibly guarantee equitable power-sharing and fair play is associated with poor public goods provision. It is, therefore, also associated with inadequate supply and poor quality of government statistics insofar as the latter are public goods.³⁶ The disruption of democratic processes can also affect the availability and quality of government data by restricting data collection. Surveyors might be unable to access sites due to disturbed conditions.³⁷ While the weakening of democracy affects statistics, democratic processes can themselves be affected by the quality of statistics. The use of flawed data in policymaking affects the relationship between the government and the beneficiaries of the flawed data, on the one hand, and aggrieved communities, on the other. Conflicts over redistribution on the basis of census are cases in point. In Nagaland, a state in north-eastern India, which will be explored in greater detail in the following chapters, flawed headcounts affected the sharing of public resources and divided the population into warring camps.³⁸ So, *democracy* and *data* deficits are also interrelated.

Data deficit is also related to development deficit. Underdevelopment limits resources available for non-essential activities such as data collection.³⁹ There is another more overt channel through which underdevelopment affects data quality. In societies with stagnant economies, where government spending apportioned among communities/regions according to reported population share and measured socio-economic disadvantage of groups plays an important role in everyday lives, statistics are manipulated to secure better representation in elected bodies and a larger share in public spending and employment. In India, the economies of insurgency-prone states such as Nagaland, located in the ethno-geographic periphery, are kept afloat by preferential federal transfers that account for more than half of the state budgets. In some of these states, communities resort to competitive manipulation of statistics. The resultant data deficit feeds back into development planning.⁴⁰ Thus, there is a bidirectional relationship between *data* and *development* deficits.

To sum up, *development*, *democracy* and *data* deficits are interrelated and as a result the quality of data is both affected by and affects political and economic processes. Statistics impact politics and policymaking, even as the latter intervene at various stages of production, dissemination and consumption of statistics. In other words, statistics are 'an integral part of the economic and social world, which they

12 *Numbers in India's Periphery*

seek to describe', and, consequently, descriptions of statistics 'should be integrated within the wider history of the society that produces them' (Tooze 2003: 3; see also Diaz-Bone and Didier 2016: 9; Alonso and Starr 1987). However, the quality of data, let alone the mutually constitutive relationship among statistics, politics and economy, has not received sufficient scholarly attention. Demographers Guilмото and Irudaya Rajan (2013: 69) note, 'In case of discrepancies, it is often found easier to dismiss statistical data than engage them, leading researchers to either credulously accept statistics or indiscriminately ignore them.' Likewise, in political science, researchers 'make heavy use of census statistics but have given scant attention to the politics behind the production of those statistics ...' (Prewitt 2010: 237). Given the interconnectedness of statistics and a variety of social phenomena, 'the relative paucity of studies of how statistics became what they are today is somewhat surprising' for historian Woolf (1989: 588). Others note the neglect of government statistics in academic research and university curriculum (Gal and Ograjenšek 2017: 80).

Economist Wade (2012: 17) points out that '[t]he sub-field of the political economy of statistics is notable for its absence', which is odd given that 'National and international statistical offices always operate in the tension between professional standards of objectivity and political insistence on certain results' (ibid.: 18). Srinivasan (1994: 4) observes that

it would appear that researchers either are not aware of or, worse still, have chosen to ignore the fact that the published data, national and international, suffer from serious conceptual problems, measurement biases and errors, and lack of comparability over time within countries and across countries at a point in time. Of course this fact is neither new nor of recent origin. After all, three decades ago Morgenstern (1963) published his classic on the accuracy of economic and social data.

Commenting on the paucity of 'studies focusing on data adequacy and quality in India', Shetty (2012: 41, 43) suggests that 'collection of accurate statistics has a low priority in policymaking today and the intellectual community which studies India's economic problems also shows no concern for the deteriorating quality of the Indian database, which they otherwise studiously use for economic analysis or for various econometric exercises'.⁴¹ Jerven (2013: xiv) too notes the 'surprising gap between knowing innately that these numbers cannot be good and an unwillingness [in economics] to study how bad they are' and adds that '[t]he scholars who are best equipped to analyze the validity and reliability of economic statistics are often data users themselves and are thus reluctant to undermine the datasets that are the bread and butter of scholarly work' (ibid.: 8). Philipsen (2015) highlights moral hazard on part of economists who avoid questioning the validity