



# Statistical Glossary

This glossary provides straightforward explanations of the terms that guide The Bahamas' national statistics, enabling policymakers, students, researchers and business leaders to interpret data with clarity and confidence.

## A

### **Administrative data**

Data collected primarily for administrative or operational purposes (e.g. tax, social security, education records) that are later used to produce statistics.

### **Age-specific fertility rate (ASFR)**

Number of live births to women in a specific age group per 1,000 women in that age group in a given year.

### **Age-specific mortality rate**

Number of deaths in a specific age group per 1,000 (or 100,000) people in that age group in a given year.

### **Aggregate**

A value derived by summing or otherwise combining detailed data for groups, categories, or geographic areas.

### **Annual growth rate**

The percentage change in a variable (such as GDP, population, CPI) between one year and the next.

### **Asset (economic)**

A store of value representing ownership of something that can provide future economic benefits, such as buildings, machinery, or financial instruments.

## B

### **Balance of payments (BoP)**

A systematic record of all economic transactions between residents of an economy and the rest of the world over a period of time.

### **Base year (reference year)**

The year chosen as a benchmark for index numbers or constant-price series (e.g. CPI base year 2019 = 100).

### **Benchmarking**

Adjusting or aligning short-term or survey-based estimates to more reliable or comprehensive reference data, often from censuses or annual accounts.

### **Big data**

Large, complex, and often real-time datasets generated by digital activities (e.g. mobile records, web traffic) that may be used for statistics.

### **Birth (live birth)**

The complete expulsion or extraction of a baby from the mother, showing any sign of life, regardless of gestational age.

## C

### **Census**

A complete count of all units in a population (e.g. all persons, all dwellings) at a specified point in time, usually conducted at regular intervals.

### **Coefficient of variation (CV)**

A measure of relative variability, calculated as the standard error divided by the estimate, often expressed as a percentage.

### **Consumer Price Index (CPI)**

An index measuring the average change over time in the prices paid by households for a fixed basket of goods and services.

### **Confidentiality (statistical)**

The obligation to protect information so that data about individual persons, households, or businesses cannot be disclosed or inferred.

### **Confidence interval**

A range of values within which the true population parameter is expected to lie with a given probability (e.g. 95%).

**Cross-section data**

Data collected for many units (e.g. persons, firms, islands) at a single point in time.

## D

**Data editing**

Activities to detect and correct errors or inconsistencies in collected data before estimation and dissemination.

**Data integration**

Combining data from different sources (surveys, administrative data, big data) to create more complete or higher-quality statistics.

**Data validation**

Checking collected data against rules, ranges, or external sources to ensure accuracy and consistency.

**Deflator (price deflator)**

A price index used to convert current-price values into constant-price values (e.g. GDP deflator).

**De facto population**

All persons physically present in the country on the reference date, regardless of usual residence or nationality.

**De jure population**

All persons who usually reside in the country, regardless of where they are on the reference date.

## E

**Economically active population (labour force)**

All persons of working age who are either employed or unemployed but actively seeking and available for work.

**Employment**

The number of persons who, during a reference period, performed work for pay, profit, or family gain.

**Error (sampling error)**

The difference between an estimate based on a sample and the true but unknown population value, arising because only part of the population is observed.

**Estimation**

The process of using sample data and statistical methods to infer values for an entire population.

**Expenditure approach (to GDP)**

A method of calculating GDP by summing final expenditures on goods and services: consumption, investment, government expenditure, and net exports.

**F****Final consumption expenditure**

Spending by households, non-profit institutions, and government on goods and services used for direct satisfaction of individual or collective needs.

**Frame (sampling frame)**

A list or other representation of all units in the target population from which a sample is drawn.

**Frequency distribution**

A table or graph showing how often each value or range of values occurs in a dataset.

**G****General government**

All institutional units that provide non-market services, redistribute income and wealth, and are mainly financed by compulsory payments (taxes).

**Geospatial data**

Data that include geographic coordinates or shapes, allowing statistics to be mapped to specific locations.

**Gross Domestic Product (GDP)**

The total value of goods and services produced in an economy during a given period, minus the value of intermediate consumption.

**Gross National Income (GNI)**

GDP plus net primary income from abroad (income received by residents from non-residents minus similar payments to non-residents).

**H****Household**

A group of persons, or a single person, living together and sharing living arrangements as a unit.

**Household consumption expenditure**

The value of goods and services purchased or received in kind by households for their own use.

**Household survey**

A survey that uses households as the primary sampling unit and collects social, demographic, and economic information about their members.

**I****Imputation**

The process of replacing missing or inconsistent data with substituted values based on other information, to complete a dataset.

**Index number**

A statistic that expresses the level of a variable relative to a base period (e.g. 2019 = 100).

**Informal sector**

Production units that are typically small in scale, unregistered, and operating with low levels of organisation and formality.

**International standard (statistical)**

An agreed international framework (such as the System of National Accounts, SNA) that provides definitions and classifications for comparable statistics.

**J****Job**

An explicit or implicit contract in which a person works for a unit in exchange for compensation.

**Job vacancy**

A paid job position that is newly created, unoccupied, or about to become vacant, and for which the employer is taking active steps to find a suitable candidate.

**K****Key indicator**

A statistic selected to summarise or monitor important aspects of social, economic, or environmental conditions (e.g. unemployment rate, poverty rate).

**Kurtosis**

A measure of the “tailedness” of a distribution, describing how concentrated values are in the centre versus the tails.

## L

### **Labour force participation rate**

The labour force (employed plus unemployed) as a percentage of the working-age population.

### **Life expectancy at birth**

The average number of years a newborn is expected to live if prevailing mortality patterns continue throughout life.

### **Line graph**

A chart type where data points are connected by lines to show trends over time.

## M

### **Margin of error**

The range around an estimate within which the true population value is expected to fall with a stated level of confidence.

### **Metadata**

Data that describe other data, such as definitions, classifications, methods, sources, and quality information for a statistic.

### **Microdata**

Record-level data about individual units (e.g. persons, households, businesses), usually anonymised to protect confidentiality.

### **Mode**

The value that occurs most frequently in a dataset.

### **Monitoring and evaluation (M&E)**

Systematic collection and analysis of data to track progress and assess results of programmes, policies, or strategies.

## N

### **National accounts**

A coherent, integrated set of macroeconomic accounts that describe the economic activities of a country, following the System of National Accounts (SNA).

### **National Statistical Office (NSO)**

The main national agency responsible for producing and coordinating official statistics in a country.

**Non-response**

Failure to obtain information from some sampled units, either entirely (unit non-response) or for some questions (item non-response).

**Non-sampling error**

Errors not related to sampling, such as mistakes in data collection, processing, or coverage.

## O

**Official statistics**

Statistics produced and disseminated by the national statistical system, following professional standards of quality, impartiality, and confidentiality, to serve the public good.

**Outlier**

An observation that is unusually large or small compared with most of the data and may strongly influence statistical results.

**Output (economic)**

The value of goods and services produced by an institutional unit or industry during a given period.

## P

**Panel data (longitudinal data)**

Data collected from the same units (e.g. households, firms) at multiple points in time.

**Population (statistical)**

The complete set of units (persons, households, businesses, etc.) about which information is sought.

**Population estimate**

A measure of the size and characteristics of a population at a given time, produced using censuses and other data sources.

**Poverty line**

The threshold level of income or consumption used to distinguish between poor and non-poor individuals or households.

**Prevalence rate**

The proportion of a population that has a particular characteristic or condition at a specific point (point prevalence) or period (period prevalence).

## Q

### **Quality (of statistics)**

A multi-dimensional concept including relevance, accuracy, timeliness, comparability, coherence, accessibility, and clarity.

### **Quartile**

One of three cut points that divide a distribution into four equal parts.

## R

### **Rate**

A measure describing the frequency of an event relative to a population at risk, often per 1,000 or 100,000 people.

### **Reference period**

The time period to which the data relate (e.g. week, month, year).

### **Register (statistical register)**

A continuously updated database of units (such as businesses or dwellings) used as a sampling frame and source of information.

### **Regression analysis**

A statistical technique used to examine the relationship between one dependent variable and one or more independent variables.

### **Response rate**

The proportion of sampled units that provided usable information.

## S

### **Sample**

A subset of units selected from a population for measurement.

### **Sample survey**

A survey that collects data from a sample rather than the full population, and uses statistical methods to generate estimates.



**Sampling error**

The error arising because a sample rather than the full population is observed.

**Sampling frame**

See “Frame”.

**Seasonal adjustment**

A statistical technique that removes the effects of regular seasonal patterns from a time series to reveal underlying trends.

**Social statistics**

Statistics on social conditions and wellbeing, such as education, health, crime, and social protection.

**Standard deviation**

A measure of the spread of a distribution indicating how far values typically deviate from the mean.

**Standardised rate**

A rate adjusted to a standard population structure to allow comparison across regions or over time.

## T

**Tabulation**

The process of organising data into tables, showing categories, counts, or other statistics.

**Target population**

The entire group of units about which information is desired and to which survey results are intended to apply.

**Time series**

Data for the same variable recorded at successive points in time.

**Timeliness**

The length of time between the reference period and the publication of statistics.

## U

**Unemployed (standard definition)**

Persons without work during the reference period, currently available for work, and actively seeking work.

**Unemployment rate**

The number of unemployed persons as a percentage of the labour force.

**Unit (statistical unit)**

The entity about which information is collected, such as a person, household, enterprise, or establishment.

## V

**Value added (gross value added, GVA)**

Output minus intermediate consumption, representing the value generated by production.

**Variance**

The average of the squared deviations of values from the mean; a measure of dispersion.

**Vital statistics**

Statistics derived from civil registration or similar systems on births, deaths, marriages, and other vital events.

## W

**Weight (survey weight)**

A factor assigned to each sampled unit that indicates how many population units it represents in estimation.

**Working-age population**

The population in the age range defined for labour statistics (commonly 15–64, or as specified in national practice).

## Y

**Year-on-year change**

The percentage change in a variable compared with the same period in the previous year.

**Youth unemployment rate**

Unemployment rate for persons in a specified youth age group (e.g. 15–24 years).

# Z

## **Z-score (standard score)**

A value expressing how many standard deviations an observation lies above or below the mean of the distribution.