Bias (statistics)

Statistical bias is a feature of a statistical technique or of its results whereby the expected value of the results differs from the true underlying quantitative parameter being estimated. The bias of an estimator of a parameter should not be confused with its degree of precision as the degree of precision is a measure of the sampling error. Mathematically Bias can be Defined as:

Let $\boldsymbol{T}$ be a statistic used to estimate a parameter $\boldsymbol{\theta}$. If $\boldsymbol{E(T) = \theta + bias(\theta)}$ then $\boldsymbol{bias(\theta)}$ is called the bias of the statistic $\boldsymbol{T}$, where $\boldsymbol{E(T)}$ represents the expected value of the statistics $\boldsymbol{T}$. If $\boldsymbol{bias(\theta) = 0}$, then $\boldsymbol{E(T) = \theta}$. So, $\boldsymbol{T}$ is an unbiased estimator of the true parameter, say $\boldsymbol{\theta}$ [1].

Contents

Types

A statistic is **biased** if it is calculated in such a way that it is systematically different from the population parameter being estimated. The following lists some types of biases, which can overlap.

- Selection bias involves individuals being more likely to be selected for study than others, biasing the sample. This can also be termed *Berksonian bias*.[2]
  - Spectrum bias arises from evaluating diagnostic tests on biased patient samples, leading to an overestimate of the sensitivity and specificity of the test.
- The bias of an estimator is the difference between an estimator's expected value and the true value of the parameter being estimated.
  - Omitted-variable bias is the bias that appears in estimates of parameters in regression analysis when the assumed specification omits an independent variable that should be in the model.
- In statistical hypothesis testing, a test is said to be **unbiased** if, for some alpha level (between 0 and 1), the probability the null is rejected is less than or equal to the alpha level for the entire parameter space defined by the null hypothesis, while the probability the null is rejected is greater than or equal to the alpha level for the entire parameter space defined by the alternative hypothesis.[3]
- Detection bias occurs when a phenomenon is more likely to be observed for a particular set of study subjects. For instance, the syndemic involving obesity and diabetes may mean doctors are more likely to look for diabetes in obese patients than in thinner patients, leading to an inflation in diabetes among obese patients because of skewed detection efforts.
- In educational measurement, bias is defined as "Systematic errors in test content, test administration, and/or scoring procedures that can cause some test takers to get either lower or higher scores than their true ability would merit. The source of the bias is irrelevant to the trait the test is intended to measure." [4]
- Funding bias may lead to the selection of outcomes, test samples, or test procedures that favor a study's financial sponsor.
- Reporting bias involves a skew in the availability of data, such that observations of a certain kind are more likely to be reported.
- Analytical bias arises due to the way that the results are evaluated.
- Exclusion bias arise due to the systematic exclusion of certain individuals from the study.
- Attrition bias arises due to a loss of participants e.g. loss to follow up during a study.[5]

- Recall bias arises due to differences in the accuracy or completeness of participant recollections of past events. e.g. a patient cannot recall how many cigarettes they smoked last week exactly, leading to over-estimation or under-estimation.
- Observer bias arises when the researcher subconsciously influences the experiment due to cognitive bias where judgment may alter how an experiment is carried out / how results are recorded.

See also

- Trueness
- Systematic error

References

1. Imdad Ullah, Muhammad (29 June 2012). "Bias: The Difference Between the Expected Value and True Value" (http://itfeature.com/statistics/bias-the-difference-between-the-expected-value-and-true-value). *Basic Statistics and Data Analysis*. Retrieved 27 June 2019.
2. Rothman, K.J. *et al.* (2008) *Modern Epidemiology* (Lippincott Williams & Wilkins) pp.134-137.
3. Neyman, J; Pearson, E S (1936). "Contributions to the theory of testing statistical hypotheses". *Stat. Res. Mem.* **1**: 1–37.
4. National Council on Measurement in Education http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorB Archived (https://web.archive.org/web/20170722194028/http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorB) 2017-07-22 at the Wayback Machine
5. Higgins, Julian PT; Green, Sally (March 2011). *Cochrane Handbook for Systematic Reviews of Interventions* (http://handbook.cochrane.org/chapter_8/8_4_introduction_to_sources_of_bias_in_clinical_trials.htm). The Cochrane Collaboration.

External links

- Bias: The Difference Between the Expected Value and True Value (http://itfeature.com), maintained by Muhammad Imdad Ullah.
