International Standard

ISO 24481-1

First edition
2026-08

Statistical methods for implementation of Six Sigma — Exploratory data analysis —

Part 1: General methodology

Méthodes statistiques pour la mise en œuvre de Six Sigma — Analyse exploratoire des données —

Partie 1: Méthodologie générale

Reference number
ISO 24481-1:2026(en)

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 7, *Applications of statistical methods and related techniques: transformation for excellence in process, product and services*.

A list of all parts in the ISO 24481 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user’s national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Like many other sciences and technologies, a core idea of Six Sigma is data-driven decision making, that is, exploiting the data from measurements or simulations at various points in the life cycle of a product or certain service. It is widely accepted that statistics plays a very important role in such fields. In any situation where statistics is applied, the analyst will follow a process, more or less formal, to reach findings, recommendations, and actions based on the data. There are two phases in this process:

a) exploratory data analysis (EDA);

b) confirmatory data analysis (CDA).

In 1970s, EDA was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses. Generally speaking, the emphasis in EDA is on letting data talk by themselves or hypothesis generation. In EDA efforts, the analyst searches for clues in the data that help identify theories about underlying behaviours. In contrast, the focus of CDA is hypothesis testing and inference. CDA consists of confirming these theories and behaviours. CDA follows EDA, and together they make up statistical modelling during data analysis.

Since the rapid development of big data and data science, there is a clear need to evolve the mechanics of Six Sigma both to accommodate the greater availability of data and to address the fact that, historically, approaches to analysing data were overly concerned with hypothesis testing, to the detriment of the hypothesis generation and discovery needed for improvement. Six Sigma teams often find warehouses of data that are relevant to their efforts. Discovery is largely supported by the generation of hypotheses—conjectures about relationships and causality. Today's Six Sigma teams, and data analysts in the business world in general, are often trained with a heavy emphasis on hypothesis testing, with comparatively little emphasis given to hypothesis generation and discovery. They are often hampered in their problem-solving and improvement efforts by the inability to exploit exploratory methods, which could enable them to make more rapid progress, often with less effort.

Though the concept of EDA has evolved and is still evolving, the main idea of it is unchanged. In most data analysis process, EDA is an approach for data analysis that employs a variety of techniques (mostly graphical) to

— maximize insight into a data set;

— identify data issues (e.g. outliers, missing values, anomalies);

— uncover underlying structure, patterns and relationship;

— extract important variables;

— determine optimal factor settings;

— enhance problem-solving, root cause analysis, and continuous improvement efforts in process optimization.

In recent times, we have seen incredible advances in visualization methods which is very important and useful to EDA, supported by phenomenal increases in computing power. Some software packages have incorporated such methods. Other software tools are continuously developed, with richer graphical functionalities. These techniques and approaches should be further promoted within Six Sigma projects to fully leverage the benefits of EDA.

Statistical methods for implementation of Six Sigma — Exploratory data analysis —

Part 1: General methodology

1 Scope

This document describes the concept and general principles of EDA for categorical and numerical data. It also provides some guidelines for conducting EDA and its place within Six Sigma projects. This document focuses on the graphical tools of EDA.

It is applicable to organizations using manufacturing processes as well as service and transactional processes.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
population
totality of items under consideration

[SOURCE: ISO 3534-1:2006, 1.1, modified — Notes 1, 2 and 3 deleted.]

3.2
sample
subset of a *population* (3.1) made up of one or more sampling units

[SOURCE: ISO 3534-1:2006, 1.3, modified — Notes 1 and 2 deleted.]

3.3
observed value
obtained value of a property associated with one member of a *sample* (3.2)

[SOURCE: ISO 3534-1:2006, 1.4, modified — Notes 1 and 2 deleted.]