

Data Cleaning Audit Checklist

Create an analysis-ready dataset without losing the audit trail.

1. Provenance and structure

- ☐ Record the source, extraction date, inclusion window, owner, and data version.
- ☐ Confirm the row unit, expected row count, unique identifier, and permitted duplicate structure.
- ☐ Preserve the raw source as read-only and perform cleaning in reproducible code or logged steps.

2. Variables and values

- ☐ Match every field to the data dictionary, measurement unit, type, coding scheme, and valid range.
- ☐ Review missing-value codes, impossible values, category spelling, dates, and cross-field consistency.
- ☐ Investigate outliers in context before correcting, excluding, transforming, or retaining them.

3. Transformations

Transformation	Rule or rationale	Rows affected	Checked by
Recode			
Derived variable			
Exclusion			
Imputation			

4. Final validation

- ☐ Recalculate counts and summaries after cleaning and reconcile them with the source.
- ☐ Confirm analysis variables, labels, reference categories, and units with the analysis plan.
- ☐ Save the final dataset, cleaning log, code, dictionary, and unresolved issue list together.

Use this resource with the related DataStatPro guide. Adapt every decision to the approved protocol, data structure, and destination requirements.