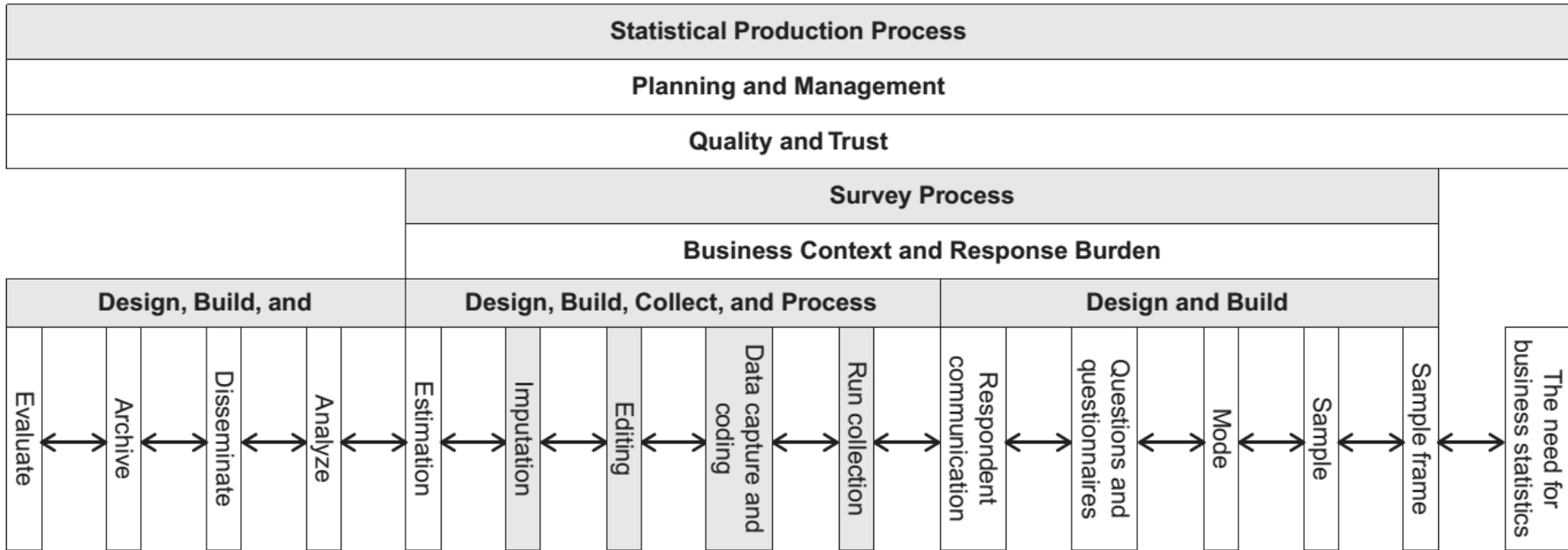# Data Processing

Labour Cost Statistics

6-7 March 2017

Jakarta- Indonesia

Presentation plan

- Data capturing

- Editing

- Turkstat practices

Source: Snijkers, Haraldsen, Jones-2013

Initial Data Capture Using Different Modes of Data Collection

Data Collection Mode	Initial Data Capture by	Initial Data Capture Potential Sources of Error
Web questionnaires	Respondents	Keying
Spreadsheets	Respondents	Keying
TDE	Respondents	Keying
IVR	IVR recognition software	Software recognition
CATI	Interviewers	Mishearing and keying
CAPI	Interviewers	Mishearing and keying
Paper self-completion	Respondents	Writing

Source: Snijkers, Haraldsen, Jones-2013

Mode of Data Collection and Data Capture Processing Methods

Data Collection Mode	Data Capture Processing Method
Web Spreadsheets TDE CATI CAPI	Data are transferred to a server and ideally are loaded into an integrated data management system (e.g., a data warehouse); raw data are electronically combined into a datafile
IVR	Data are transferred to a server; verbal recordings are translated to alpha characters and numerical values using speech recognition; translated data are electronically combined into a datafile
Paper questionnaires	Manual keying, electronic scanning, or combination thereof

Initial Data Capture and Data Capturing Processing Potential Sources of Error

Data Collection Mode	Initial Data Capture by	Initial Data Capture Potential Source(s) of Error	Data Capture Processing by	Data Capture Processing Potential Source(s) of Error
Web, spreadsheets, TDE	Respondents	Keying	Electronic integration	System specification
CATI, CAPI	Interviewers	Mishearing and keying	Electronic integration	System specification
IVR	Recognition software	Software recognition	Translating verbal responses to characters and values; and electronic integration	Translation and system specification
Paper self-completion	Respondents	Writing	Keying, scanning, or a combination of keying and scanning	Keying, scanning, system specification
			Electronic integration	

Editing

Error Types	Detection	Treatment
Missing values	Logical edits	If missing: impute
	Validity and range edits	If not missing: use special character to identify it
	Consistency edits	Flag: detection and treatment
	Fellegi–Holt edits	
Systematic errors	Expert subject matter knowledge	Flag: detection and treatment
	Statistical edits: ratio edits	
Random errors	Logical edits	Use deterministic imputation
	Validity and range edits	Use Fellegi–Holt principles to impute correct values
	Consistency edits	Flag: detection and treatment
Influential errors	Selective editing	Impute within allowed range
	Score functions	Flag: detection and treatment
	Graphical techniques	
Outliers	Robust univariate methods	If value is correct, identify as such, and use robust methods in estimation procedures; if incorrect, either impute within allowed range or follow up with respondent
	Hidiroglou–Berthelot methods	
	Regression models	
	Multivariate methods	
	Graphical techniques	
		Flag: detection and treatment

Editing

- Micro Editing

- Macro Editing
  - Aggregate Method
  - Top–Down Method
  - Graphical Methods

- Imputation
  - Logical (or Deductive) Imputation
  - Mean-Value Imputation
  - Historical Imputation
  - Nearest -Neighbor Imputation

TurkStat Practices