Secondary Data and Its Sources in the Agriculture and Allied Sector

K. Naga Latha[1*], Tadaveni Anitha[1], Tadepalli Yamini[1] and V. Mathuabirami[1]

[1]Assistant Professor, School of Agriculture, Kaveri University, Gowraram, Telangana-502279

*Corresponding author mail id: nagalatha.k@kaveriuniversity.edu.in

Data serves as the foundational raw material in statistical analysis, transformed through statistical methods into meaningful information and insights about specific phenomena. While raw data alone such as area, production, and yield figures for paddy offers no interpretive value, statistics unlock patterns like growth rates, temporal shifts, or productivity trends, highlighting the fundamental distinction between mere data collection and statistical inference. For agricultural researchers and educators, this differentiation is critical when designing experiments or preparing teaching materials, as it underscores the need for analytical tools like ANOVA or regression to extract actionable knowledge from datasets.

Primary data involves firsthand collection by the researcher through interviews, questionnaires, focus groups, observation, the examination of primary sources such as writings or speeches, or a variety of other such methods (Vartanian, 2010). However, these methods demanding substantial manpower, time, and financial resources. In contrast, secondary data comprises information already gathered and published by others, accessible from government portals, reports, or databases at minimal cost and effort. The acquisition of secondary data transcends statistics, embodying multidisciplinary appeal across economics, sociology, and agriculture, where pre-existing datasets are repurposed for novel analyses (Daas and Beukenhorst, 2008). This aligns with secondary research methodologies, which repurpose archived data for objectives distinct from their original intent (Golden, 1976 & David et al., 1993).

Secondary Data in Agricultural Statistics

Secondary data in agriculture draws from diverse repositories, including government agencies, international bodies, industry reports, academic journals, and commercial publications. Government sources such as India's Directorate of Economics and Statistics (DES), National Sample Survey Office (NSSO), or global entities like the Food and

Agriculture Organization (FAO) provide authoritative datasets on crop statistics, land use, trade, and prices. Industry reports from export promotion councils (e.g., APEDA, MPEDA) offer trade volumes, while archived datasets from the Economic and Political Weekly (EPW) database furnish time-series on productivity and inputs. Traditional outlets like books, peer-reviewed journals, and newspapers supplement with contextual narratives. These sources deliver rapid, cost-effective responses to research queries, enabling comparisons across regions, crops, or years without fieldwork.

Advantages of secondary data

Secondary data helps to define a research problem, formulate research questions and hypotheses, and select a research design (Clark, 2014). Secondary data streamlines research by clarifying problem definitions through preliminary exploration of existing trends, such as paddy yield growth under varying irrigation regimes. Its low cost and time investment allows educators to quickly assemble datasets for R-based visualizations or classroom ANOVA demonstrations, bypassing the logistics of primary surveys.

- **Problem Formulation Aid:** Initial access to historical crop production data from DES or FAO helps refine hypotheses, e.g., identifying yield gaps for factorial experiments.
- **Economic Efficiency:** No need for extensive sampling; download CSV files from NSSO for land holdings analysis in minutes.
- **Accuracy Potential:** Official government portals ensure standardized, verified metrics, outperforming ad-hoc primary efforts.
- **Comparative Power:** Facilitates benchmarking, like contrasting Indian wheat yields against USDA global averages for teaching competitiveness.

Disadvantages of secondary data

Despite benefits, secondary data poses challenges that demand rigorous evaluation, particularly in agricultural contexts where unit inconsistencies or temporal lags can skew interpretations.

- **Accessibility Barriers:** Non-open-source data, such as proprietary EPW time-series, requires subscriptions; prioritize public portals like FAO's FAOSTAT.
- **Reliability Variability:** Third-party aggregators may introduce errors; cross-validate with primary government sites (e.g., DES over blogs).
- **Obsolescence Risks:** Outdated figures (pre-2025) limit forecasts; use latest releases from DOES for current kharif/rabi estimates and extrapolate cautiously via ARIMA models.

- **Unit and Definitional Incompatibilities:** Paddy area in acres (older NSSO rounds) versus hectares (FAO) necessitates conversions (1 hectare = 2.471 acres); standardize in R scripts prior to analysis.

Sources of secondary data in Agriculture:

Organizations worldwide and in India serve as vital repositories of secondary data for agricultural statistics, enabling researchers, educators, and policymakers to conduct robust analyses such as ANOVA, trend forecasting, and structural equation modeling without primary data collection. These sources provide categorized datasets on crop production, land utilization, trade dynamics, and market prices, which are indispensable for book chapters on agricultural statistics, experimental design teaching, and data visualization in R.

1.Crop Production Statistics

Crop statistics encompass data on sown area, harvested area, yield per hectare, production volumes, and input usage (e.g., fertilizers, irrigation), forming the backbone for design of experiments (e.g., CRD, RBD) and productivity assessments. This data is crucial for understanding temporal trends, climate impacts, and varietal performance, allowing educators to demonstrate ggplot2 visualizations of yield gaps or ANOVA tables in classroom settings. Access these through official portals for downloadable CSV/Excel files, often with historical series spanning decades.

FAO (Food and Agriculture Organization): Offers global databases like FAOSTAT, covering production quantities, yields, and land under major crops (cereals, pulses, oilseeds, roots/tubers) across 245 countries since 1961. Includes fertilizer consumption, irrigation coverage, and machinery use; ideal for cross-country comparisons in teaching global agriculture.

DOES (Directorate of Economics & Statistics, Ministry of Agriculture & Farmers Welfare, India): Publishes annual "Agricultural Statistics at a Glance" with India-specific kharif/rabi crop estimates, area-yield-production triads, cost of cultivation (COC) surveys for 27 crops, and input-output ratios. Essential for Indian factorial experiment analyses and policy evaluations.

USDA (United States Department of Agriculture): Provides detailed datasets via Economic Research Service (ERS) on principal field crops (corn, wheat, soybeans, cotton), livestock products, and long-term projections (e.g., WASDE reports). Supports econometric modeling of supply-demand balances.

2. Land Use and Holdings Statistics

Land statistics detail classifications (net sown, fallow, cultivable waste), irrigation intensity, cropping patterns, and operational holdings by size/class, critical for resource

allocation studies, soil fertility modeling, and impact assessments of land reforms. In teaching, these datasets facilitate R scripts for spatial analysis or simulations of land productivity under different experimental layouts. Data is typically available in tabular formats with state/district breakdowns.

NSSO (National Sample Survey Office): Conducts periodic surveys (e.g., Situation Assessment of Agricultural Households) yielding data on land holdings distribution (marginal to large), tenancy patterns, leased-in/out areas, and irrigated vs. rainfed parcels across rural India.

DOES (Ministry of Agriculture, India): Tracks land utilization statistics annually, including gross cropped area, net irrigated area, cropping intensity, and fallow lands by state. Complements crop data for comprehensive land productivity indices.

EPW Research Foundation Database: Curates time-series on land use patterns, fragmentation trends, and regional disparities, drawn from official sources; useful for longitudinal studies.

3.Export-Import Trade Data

Agri-trade data includes export/import volumes (quantity/value in USD/Rupees), country/port-wise breakdowns, product schedules (HS codes), and policy metrics (tariffs, subsidies), pivotal for trade balance analysis, competitiveness indices, and visualization of export growth curves. For book chapters, this supports case studies on globalization's impact on Indian agriculture, with data exportable for R-based heatmaps or regression models.

APEDA (Agricultural & Processed Food Products Export Development Authority): Focuses on 50+ scheduled products (fruits, vegetables, cereals, spices, processed foods and meat); provides annual export figures, market intelligence reports, buyer directories, and RUPIA/Dollar conversions since inception.

MPEDA (Marine Products Export Development Authority): Details marine exports (shrimp, fish, squid) by item, quantity/value, destination markets (USA, EU, Japan), and port-wise shipments from 1995; includes traceability and sustainability metrics.

WTO (World Trade Organization): Global agri-trade profiles with tariff bindings, domestic support notifications, export subsidies, and dispute data; enables analysis of AoA (Agreement on Agriculture) compliance.

USDA: International trade datasets for grains, oilseeds, meats; includes monthly updates and bilateral flows.

4. Agricultural Commodity Prices

Price datasets cover wholesale/retail quotes, arrivals (quantity traded), modal prices, support prices (MSP), and indices (e.g., WPI for agri), enabling volatility modeling, parity

ratio calculations, and forecasting via ARIMA in R. These are invaluable for teaching market efficiency, price transmission, and risk management in agricultural economics chapters.

Directorate of Marketing & Inspection (DMI): AGMARKNET portal offers real-time daily/weekly data on arrivals, minimum/average/maximum prices for 200+ commodities across 7,000+ markets; includes grading standards and trend reports.

DOES (Ministry of Agriculture): Historical MSP announcements, season-wise harvest prices, input-output price ratios, and state-level wholesale prices linked to crop estimates.

EPW Database: Long-term commodity price series, wholesale price indices (WPI), and inter-crop price ratios for econometric analyses.

References:

Vartanian, T. P. (2010). *Secondary data analysis*. Oxford University Press.

Clark, G. (2014). Data sources for studying agriculture. In *The geography of agriculture in developed market economies* (pp. 32-55). Routledge.

Daas, P., & Beukenhorst, D. (2008). *Databronnen van het CBS: Primaire en secundaire bronnen (Statistics Netherlands data sources: Primary and secondary sources)*. Internal report, Statistics Netherlands, Heerlen.

Golden, B. (1976). Shortest-path algorithms: A comparison. *Operations Research*, *24*(6), 1164-1168.

David, L. P., Slyz, A., Jones, C., Forman, W., Vrtilek, S. D., & Arnaud, K. A. (1993). A catalog of intracluster gas temperatures. *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 412, no. 2, p. 479-488.*, *412*, 479-488.

https://www.india.gov.in/website-directorate-economics-and-statistics

https://agriculture.institute/qualitative-quantitative-analysis-for-agribusiness/types-examples-secondary-data-sources/

http://www.ers.usda.gov/data-products

https://mpeda.gov.in

https://apeda.gov.in/

https://agriwelfare.gov.in

https://services.india.gov.in/service/detail/agmarknet-portal

https://fdc.nal.usda.gov