# Research Assistant Skills Test – Data Handling and Documentation

This assessment evaluates your ability to clean, merge, and summarize administrative and survey data using any appropriate data-analysis platform (e.g., Stata, R, Python, or others). Your performance will be judged on accuracy, documentation quality, and reproducibility.

## General Instructions

1. You have received two datasets: payments.dta and survey.dta.

2. You may use any software platform for analysis (Stata, R, Python, etc.).

3. Your code must use global variables or dynamic file paths so it runs on any computer without edits.

4. All code must be clearly annotated—explain each step and any assumptions. Unannotated code will be penalized.

5. Submit only: one script (.do/.R/.py), one log/output file, and a short summary note (readme.txt or .pdf, ≤150 words).

6. Do not share code, outputs, or datasets with anyone. Collaboration or public sharing disqualifies your submission.

7. Expected completion time: 2 days. Submit on Monday 27th October, 10 AM Max.

**Do NOT share the datasets with anyone.**

## Data Description

payments.dta

|  |  |
| --- | --- |
| Variable | Description |
| division | Administrative division (e.g., Lahore) |
| districtname | District name |
| ratingarea\_name | Tax rating area within the district |
| circle\_name | Tax circle name |
| pin | Unique property identifier |
| street\_address | Property address |
| total\_property\_tax | Total assessed property tax |
| payable\_property\_tax | Payable tax after adjustments |
| exempted\_amount | Exempted portion of tax liability |
| remission | Waiver or remission amount |
| netdemand | Final tax demand after remission/exemption |
| year | Fiscal year (e.g., “2022-2023”) |
| totalrecovery | Total payment recovered for that year |
| bank\_code | Payment channel or bank code |
| receipttype | Type of payment record (“Current” / “Combine”) |
| region | Administrative region code |
| payment\_date | Date and time of payment |

## Tasks

### Task 1 – Data Import and Cleaning

Clean payments.dta: handle placeholder/missing codes (e.g., 97, 98, 99, –90, –99, NULL); parse fiscal years; and create derived variables: arrears\_prev (1 if receipttype = “Combine”), paid\_share (totalrecovery/netdemand), and paid\_any (1 if totalrecovery > 0). Document all checks, outliers, and data issues found.

### Task 2 – Summary Statistics

For fiscal year 2022–2023, report: total properties, share paid (paid\_any), mean paid\_share, mean netdemand, and mean totalrecovery. Provide breakdowns by region (and circle\_name if feasible).

### Task 3 – Payment Timeliness

Assume due\_date = 30 September of the fiscal year end. Compute payment delay = payment\_date – due\_date; summarize mean/median/10th/90th percentiles; and produce a histogram or bar chart of payment timing.

### Task 4 – Survey Data Analysis (with Attitudes & Preferences Module)

Use survey.dta to produce core survey metrics and an extended module on attitudes and reform preferences. Clearly annotate all steps and state assumptions. Keep outputs concise.

**A) Core survey metrics**

• Use status\_survey (or equivalent flag) to report overall response rate and response rate by treatment (treat\_arm).

• Assess treatment balance on: property land area, built area, number of storeys, age, gender, and education.

• Construct an Asset Ownership Index using PCA. List variables included and justify choices.

**B) Attitudes & Preferences Module (5 parts max)**

• Tax Knowledge & Misperceptions: Build a Tax Knowledge Index (share correct on knowledge/ATR items). Compare across education, asset proxies, and gender.

• Trust & Institutional Confidence: Create a Trust Scale (tax staff, government, citizens). Cross-tabulate by knowledge quartiles and neighborhood type (planned vs unplanned).

• Fairness & Distributional Beliefs: Construct a Fairness-Perception Index (inequality, fair share for rich/middle). Relate to support for progressive taxation.

• Reform Preferences & Expected Effects: Summarize expected impacts of raising top rates vs broad-base increases on activity, investment, mobility. Identify traits associated with support for progressive schedules.

• Policy Implications: Synthesize a short citizen typology and provide 2–3 actionable recommendations for the 2025 property-tax reform communication and sequencing.

### Task 5 – Merge Payments and Survey Data

Merge payments.dta and survey.dta using the unique identifier (pin). Compute treatment effects on: paid\_share (behavior), s4\_q6 and s4\_q18 (beliefs). In ≤150 words, summarize differences between stated beliefs and behavior, and discuss implications.

## Self-Assessment

On a scale of 1–5, indicate how much you relied on ChatGPT or similar tools: 1 = minimal help (basic syntax or debugging); 5 = copied substantial code blocks. Honest self-assessment will not affect selection; it informs onboarding.

## Bonus Task (Optional, +10 points)

Create a GitHub account and email your GitHub address to ahsan.farooqui@ideaspak.org. Upload your work to a repository, create a separate branch for your submission, and create a pull request there.