# Data and Summary Stats Section Rubric

**What to Submit**

- Keep in mind that this is the Data section that you will be dropping straight into your finished paper, with little or no editing.
- **Title**
    - **Your title should be your approved research question**
- **Written Section (about 1 page, give or take):**
    - Where your data comes from (source, coverage, unit of observation, years available).
    - Any steps you had to take to compile or merge datasets.
    - Basic cleaning decisions (e.g., dropping missing values, winsorizing, excluding observations).
    - What the variables in your summary stats table represent.
    - Any surprising, unusual, or important facts about the data (e.g., trends, large variation, unexpected values).
    - Write this in full sentences and paragraphs.
- **Summary Statistics Table:**
    - At minimum, include the mean and standard deviation of your key variables.
    - Include the number of observations.
    - Use clear variable names or labels.
    - Format the table so it is clean and professional (LaTeX is welcome but not required). In economics papers, the style is very plain.
    - The table should be as self-contained as possible. Any notes concerning key cleaning decisions or exclusions should be in a subscript beneath the table. Those points should be repeated in the text, with a full explanation to justify your decision

- This table should reflect the data AS YOU WILL USE IT for your main regressions. That is, if you are going to modify the dataset or exclude observations, the table should reflect the data after those changes.
- Look at the example table for style guidance

Table 1: Elementary School Data: Summary Statistics

| | 2000-01 thru 2018-19 | 2000-01 thru 2006-07 | 2007-08 thru 2018-19 |
|---|---|---|---|
| Mean Total Enrollment | 466 (216) | 456 (225) | 472 (211) |
| Mean White Share | 0.581 (0.332) | 0.630 (0.335) | 0.553 (0.327) |
| Mean Black Share | 0.140 (0.224) | 0.149 (0.240) | 0.134 (0.214) |
| Mean Hispanic Share | 0.204 (0.263) | 0.168 (0.251) | 0.225 (0.268) |
| Mean Asian Share | 0.042 (0.089) | 0.039 (0.087) | 0.044 (0.089) |
| Mean Other Race Share | 0.075 (0.117) | 0.053 (0.107) | 0.089 (0.121) |
| Mean Two Or More Races Share | 0.018 (0.032) | 0.000 (0.000) | 0.029 (0.037) |
| Mean Share Free Lunch | 0.419 (0.265) | 0.366 (0.249) | 0.449 (0.269) |
| % Rural | 0.243 (0.429) | 0.264 (0.441) | 0.230 (0.421) |
| % Small Town | 0.093 (0.290) | 0.091 (0.288) | 0.094 (0.291) |
| % Large Town | 0.023 (0.150) | 0.018 (0.132) | 0.026 (0.159) |
| % Small Suburb | 0.012 (0.110) | 0.003 (0.055) | 0.018 (0.132) |
| % Midsize Suburb | 0.056 (0.230) | 0.100 (0.300) | 0.030 (0.172) |
| % Large Suburb | 0.259 (0.438) | 0.265 (0.441) | 0.255 (0.436) |
| % Small City | 0.044 (0.205) | 0.012 (0.107) | 0.063 (0.243) |
| % Midsize City | 0.080 (0.271) | 0.133 (0.339) | 0.049 (0.216) |
| % Large City | 0.097 (0.297) | 0.114 (0.318) | 0.088 (0.283) |
| Number of Schools | 54,195 | 48,193 | 48,723 |
| Number of Observations | 763,413 | 282,985 | 480,428 |

School/year observations that show a total enrollment change greater than 25% are excluded, as are observations from districts that do not guarantee seats in local elementary schools. Standard errors in parentheses.

**Grading:**

**1. Written Section (50%)**

- **Title (10%):** As instructed, reflects the research question
- **Data Source and Coverage (10%):** Clearly identifies source(s), coverage, unit of observation, and years available.
- **Compilation and Cleaning (10%):** Explains merges, cleaning, and exclusions with sufficient detail..
- **Critical Observations (10%):** Notes surprising, unusual, or important features of the data.

- **Writing Quality (10%):** Section is written in full sentences and paragraphs, polished enough to drop into the final paper with minimal edits.

**2. Summary Statistics Table (30%)**

- **Completeness and Professional Formatting (10%):** Includes means, standard deviations, and number of observations for all key variables. Table is clear, clean, labeled, and follows economics style (plain, minimal, self-contained).
- **Clarity (10%):** Variable names/labels are interpretable and consistent with the text.
- **Accuracy (10%):** The table matches the data actually used (reflecting cleaning/exclusions).

**3. Data Applicability (20%):** The data as presented is sufficient to answer the research question.