**Audit Fees in SAS**

The datasets associated with this assignment contain information related to U.S. Companies that filed audited financial statements in the 2005-2014 period. Each observation in the datasets represent one firm-year. The dataset “Audit\_Info” contains information related to the firm’s audit and the dataset “Financial\_Data” contains accounting information from the firm’s 10-k. These datasets can be downloaded through Canvas.

Record the answers to any open ended questions in a Word document. **When you’ve completed all steps, copy and paste your SAS log into the same Word document.** Save the Word document and upload it through Canvas.

**Part I: Audit\_Info Dataset**

1. Import the “Audit\_Data” dataset into SAS.
2. The variable AUDIT\_FEES represents the audit fees charged to a firm in a given year. Create a new variable, “LAF”, defined as the log of audit fees.
3. The variable “AUDITOR\_CITY” is the auditor’s location. Which city had the highest average audit fees (“AUDIT\_FEES”) for the sample period? (No t-test necessary)
4. The variable AUDITOR\_FKEY is coded 1, 2, 3, or 4 for the “Big 4” auditors. Code the binary variable “BIGN” as 1 for “Big 4” auditors, and 0 otherwise.
5. Compare average audit fees earned by Big N (BIGN = 1) and non-Big N auditors (BIGN = 0). Are mean audit fees higher for Big N or non-Big N auditors? (No T-test necessary)

**Part II: Firm\_Info Dataset**

1. Import the “Firm\_Data” dataset into SAS.
2. The variable AT represents the total assets held by a firm. Create a new variable, “LAT”, defined as the log of total assets.
3. The variable SALE represents total sales for a firm-year. Create a new variable PY\_SALE defined as prior year, or lagged, sales for the same firm.
   1. After creating the new variable, be sure to code PY as “.” For observations where the lagged sales value was from a different company.
4. Define a new variable SALE\_GROWTH as (SALE minus PY\_SALE) divided by PY\_SALE.
5. Determine whether the average value of SALE\_GROWTH is significantly different from zero; perform a T-test.

**Part III: Merge**

1. Create a new dataset called “Complete\_Data” by merging the datasets “Audit\_Data” and “Firm\_Data” on the variables GVKEY and FISCAL\_YEAR\_ENDED.
2. Drop the observations for which SALE\_GROWTH is missing.
3. Run a regression that predicts the log of audit fees (LAF) as a function of: auditor size (BIGN), auditor specialization (CITY\_SPEC), restatement issuance (RESTATEMENT\_ISSUE), presence of a material weakness (MATWEAK), log of total assets (LAT), current ratio (CATA), operating cash flow (OCF), negative net income (LOSS), sales (SALE), and growth in sales (SALE\_GROWTH). Which of these variables are ***POSITIVELY AND SIGNIFCANTLY*** (p < 0.10) associated with audit fees?
4. Which observation (reference the GVKEY) has the greatest difference between predicted audit fees and actual audit fees?