

## FIN 557 Problem Set One

Due on 02/07/2023

**Instruction: format your answers to includes SAS code, brief descriptions of the code, and output results in a single PDF file. For example, copy the SAS code and descriptions in a word document and add SAS results using Screenshot tools. Organize the answers by question numbers.**

### 1. Download the dataset

Locate the database on WRDS: Get data/Compustat/Execucomp/Annual Compensation

Choose your date range: 2010-2018

Apply your company codes: search the entire database

Query variables: include the following variables

VARIABLE NAME	DEFINITION
SIC	SIC code
SICDESC	SIC Code description
TICKER	Ticker symbol
CONAME	Company name
STATE	State where the company is headquartered
CITY	City where the company is headquartered
EXECID	Executive ID number
GENDER	Executive gender
AGE	Executive's age
CEOANN	Annual CEO indicator
BECAMECEO	Date became CEO
SALARY	Salary (thousand dollars)
BONUS	Bonus (thousand dollars)
TDC1	Total compensation (salary, bonus, other annual, restricted stock grants, stock option grants, long-term incentive payouts, and all other in thousand dollars)
TDC1_PCT	TDC1 percentage change year to year (%)
YEAR	Fiscal year

Add a conditional statement to include only CEOs: CEOANN equal CEO

Select query output: output format is SAS table, and data format is mmddyy10.

Rename the dataset as ceo, create a new folder named hw1, and upload the data to the folder

Write a PROC CONTENTS procedure to list the variable attributes and information about the dataset, such as number of columns and rows

Use a PROC MEANS procedure to generate simple summary statistics for all numeric columns in the dataset

Use the dataset hw1.ceo to perform the following tasks

## **2. Exploring data with procedures**

### **2.1 List the data with a single PRINT procedure.**

Filter CEOs whose age is less than 40, and firm is in the Midwest region (North Dakota, South Dakota, Nebraska, Kansas, Missouri, Iowa, Minnesota, Wisconsin, Michigan, Illinois, Indiana, and Ohio)

Print the observations in the subset

Add a VAR statement to include only the following columns: TICKER, STATE, EXECID, CEOANN, SALARY, and TDC1\_PCT

Round values of SALARY to the nearest whole number and use a comma at the thousands position

Round TDC1\_PCT to one decimal place

### **2.2 Calculate summary statistics with a single MEANS procedure.**

Filter CEOs whose gender is female, and became CEO on or after 01/01/2010

Calculate the summary statistics for SALARY, BONUS, and TDC1 in the subset

### **2.3 Use Macro variables**

Write three %LET statements to create macro variables named SIC, GENDER, and SALARY

Filter CEOs whose SIC is 2834 (pharmaceutical preparations), gender is female, and salary is more than 1 million dollars

Use macro variables in the WHERE statement for the following two procedures

a) Print the observations in the subset with a single PRINT procedure

Add a VAR statement to include only the following columns: SIC, SICDESC, TICKER, EXECID, GENDER, and SALARY

b) Calculate summary statistics for SALARY in the subset with a single MEANS procedure.

## **3. Sort data and remove duplicate values**

### **3.1 Remove duplicate key values**

Sort hw1.ceo by YEAR and TICKER. Remove duplicate values of the column listed in the BY statement. Write the sorted data to a table named ceo1

How many observations are in ceo1?

### **3.2 Sort the data**

a) Sort ceo1 in a way such that the first row for each year represents the CEO who has the highest salary within the year. Write the sorted data to a table named ceo2

b) Print the first 10 observations in ceo2 table with a single PRINT procedure.

Add a VAR statement to include only the following columns: YEAR, TICKER, EXECID, and SALARY

### **3.3 Identifying CEO with the highest salary within each year**

Verify that the first row for each year contains the CEO who has the highest salary within the year in ceo2 table

a) Sort ceo2 table to keep the first row for each year. Write the output to a table named ceo3

b) Print the observations in ceo3 table with a single PRINT procedure.

Add a VAR statement to include only the following columns: YEAR, TICKER, EXECID, and SALARY