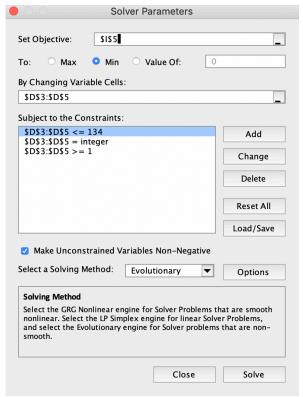
Cluster Instructions

- 1.) After downloading "Baltimore City Employee Salaries FY2018," sort and filter the data to only include Job Titles with Accounting related jobs.
- 2.) Create a Job # column to assign numerical values to each job.
 - a. In A12, put "1"
 - b. In A13, put "=A12+1"
 - i. Double click the black plus sign that appears in the bottom, right corner of the cell to apply it to the entire column.
- 3.) Find the mean of "Annual RT" and "Gross" Income columns
 - a. Mean of "Annual RT": =AVERAGE(F12:F145)
 - b. Mean of "Gross": =AVERAGE(G12:G145)
- 4.) Find the standard deviation of "Annual RT" and "Gross" Income columns.
 - a. Standard deviation of "Annual_RT": =STDEV(F12:F145)
 - b. Standard deviation of "Gross": =STDEV(G12:G145)
- 5.) Next, run a Standardize function on the "Annual RT" and "Gross" Income columns.
 - a. Utilize \$ to apply the function to the entire columns.
 - i. Z-score of "Annual RT": =STANDARDIZE(F12,F\$8,F\$9)
 - 1. Double click the black plus sign that appears in the bottom, right corner of the cell to apply it to the entire column.
 - ii. Z-score of "Gross": =STANDARDIZE(G12,G\$8,G\$9)
 - 1. Double click the black plus sign that appears in the bottom, right corner of the cell to apply it to the entire column.
- 6.) Highlight the columns all the way through "z_Gross" and title the table "Accounting_2"
- 7.) Create three Cluster IDs.
- 8.) Setting up the Solver model for Cluster Analysis
 - a. Enter trial values (integers between 1 to 134) in C3:C5 for cluster anchors
 - b. Set D5:D8 as Job #.
 - c. In E5:E8 we will utilize VLOOKUP to find the exact JOBTITLE.
 - i. Ex. =VLOOKUP(D3,Accounting 2,2)
 - d. Identify the z-score for "z_ANNUAL_RT" in the first Jobtitle using the below formula
 - i. =VLOOKUP(D3,Accounting 2,8)
 - ii. Do the same for the other two Jobs
 - e. Identify the z-score for "z Gross" in the first Jobtitle using the below formula
 - i. =VLOOKUP(D3,Accounting 2,9)
 - ii. Do the same for the other two Jobs
- 9.) Computing the squared distance
 - a. Compute the squared distance from each job title to each of the three selected cluster candidates
 - i. Ex. =SUMXMY2(\$F\$3:\$G\$3,H12:I12)
 - ii. Copy the formula for the rest of the data points.
- 10.) Compute the smallest distance for each job to the cluster anchors by using the formula

- a. =MIN(J12,K12,L12)
- b. Copy the formula for the rest of the data points
- 11.) Compute the sum of squared distance of all jobs using the formula
 - a. =SUM(M12:M145) in cell I6
- 12.) Using the solver window, find the optimal cluster anchors for the three clusters as shown below



a.