**Project-01**

**2**. It has 45 non missing zip codes. Zipcode 8888 and 9999 seems not valid zip code. So it has 43 unique zip code.

**3**. Numbers are not same for all zip code.

**4.** SAS program and the permanent SAS data set are in the folder.

5.

**SAS Code**

/\* Statifed Sampling\*/

/\* Sorting dataset by zip code\*/

**proc** **surveyselect** data=project.SFdata

out=project.stratifed

method=srs n=**5** seed=**123**;

strata zipcode;

**run**;

**proc** **freq** data=project.stratifed;

table zipcode;

**run**;

**Complete SAS Code**

\*Program name: Project1 is stored in C:\Computational Statistics\4th Quater\SAS\Project1.

Purpose: Sample selection of SFCity survey data.

Programmer: Chinki Rai

Date Written: 08th Nov 2017;

libname project "C:\Users\chink\Google Drive\Computational Statist

ics\4th Quater\SAS\Project\File2";

**PROC** **IMPORT** OUT= project.SFdata

DATAFILE= "C:\Users\chink\Google Drive\Computational Statist

ics\4th Quater\SAS\Project\File2\7041-1996-2015\_City\_Survey\_DataSF\_new.c

sv"

DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

/\* Finding unique levels \*/

title "Unquie Zipcodes";

**proc** **freq** data=project.SFdata nlevels;

table zipcode;

**run**;

/\*Total 46 zipcodes in which 1 is missing and 8888 and 9999 is not valid so 43 Unique zip code\*/

**proc** **surveyselect** data=project.SFdata

out=project.simplerandom

method=srs

sampsize=**100**;

**run**;

**proc** **freq** data=project.simplerandom;

table zipcode;

**run**;

**proc** **sort** data=project.SFdata;

by zipcode;

**run**;

/\* Statifed Sampling\*/

/\* Sorting dataset by zip code\*/

**proc** **surveyselect** data=project.SFdata

out=project.stratifed

method=srs n=**5** seed=**123**;

strata zipcode;

**run**;

**proc** **freq** data=project.stratifed;

table zipcode;

**run**;