**Project-01**

**Problem-04**

**Part- I**

1. unzipped file using unzip software and make it CSV file with the help of unzip software.
2. Code to read Zip file directly in SAS

/\* Import a csv file directly from the ZIP \*/

**data** data1987;

infile inzip(C:\Users\chink\Google Drive\Computational Statist

ics\**4**th Quater\SAS\Project\File4\**1987**\_zip.zip)

dsd ;

**run**;

SAS can not read a .bz2 file directly, so we can convert a .bz2 file to zip and can read zip file into SAS.

**Part-2**

1. Data set has 29 variables.

Character variables- Dest, Origin, UniqueCarrier, DayOfMonth, DayOfWeek Diverted, Cancelled

AirTime, CancellationCode, CarrierDelay, Dest, LateAircraftDelay, NASDelay,

Origin, SecurityDelay, TailNum, TaxiIn, TaxiOut, UniqueCarrier, WeatherDelay

Numerical Variable- ActualElapsedTime, ArrDelay, ArrTime, CRSArrTime, CRSDepTime, CRSElapsedTime, DepDelay, DepTime

**Complete SAS Code:**

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

Purpose: Reading zip file into SAS and doing some basic analysis.

Programmer: Chinki Rai

Date Written: 11th Nov 2017;

ods pdf file='C:\Users\chink\Google Drive\Computational Statist

ics\4th Quater\SAS\Project\_1\File4\surveyresult.pdf';

libname File4 "C:\Users\chink\Google Drive\Computational Statist

ics\4th Quater\SAS\Project\_1\File4";

**PROC** **IMPORT** OUT= File4.airlines

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

ics\4th Quater\SAS\Project\_1\File4\1987 \_new.csv"

DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

/\*Examining the top100 and top1000 rows of the .csv file for 1987\*/

**proc** **print** data=File4.airlines (obs=**100**);

**run**;

**proc** **print** data=File4.airlines (obs=**1000**);

**run**;

/\*1. How many variables are included in the dataset? List the Categorical variables. List the Numeric variables.\*/

**proc** **contents** data=File4.airlines;

**run**;

/\* 29 Variable in the dataset\*/

/\*2. Compute means and standard deviations for the Numeric variables for each month of 1987\*/

**proc** **means** data= File4.airlines

mean std maxdec = **3**;

var ActualElapsedTime ArrDelay ArrTime CRSArrTime CRSDepTime CRSElapsedTime

DepDelay DepTime;

class month;

**run**;

/\*3. Make tables of counts and relative frequencies for the Categorical variables for each month of 1987.\*/

**proc** **freq** data= File4.airlines;

tables (Dest Origin UniqueCarrier DayOfMonth DayOfWeek Diverted Cancelled) \* month;

**run**;

ods pdf close;

/\* Import a csv file directly from the ZIP \*/

**data** data1987;

infile inzip(C:\Users\chink\Google Drive\Computational Statist

ics\**4**th Quater\SAS\Project\File4\**1987**\_zip.zip)

dsd ;

**run**;