SAS and R Code for Introduction

**Data** task1;

input Subject Gender $ Age HSGPA CollGPA ;

datalines;

1 m 32 2.2 3.5

2 f 23 2.1 3.5

3 f 27 3.3 3.0

4 f 35 3.5 3.2

5 m 23 3.1 3.5

;

**run**;

filename survey 'c:/Documents and Settings/SMU ID/Desktop/ufldata.txt';

**Data** task2;

infile survey missover;

input Subject Gender $ Age HSGPA CollGPA HomeTown CurrRes HrsTV HrsSprt NumbNews AIDSHIV Veg $ PolAffl $ Ideology Serv AbLegal $ Affirm $ LAD $;

**run**;

PROC FSVIEW data=task2; run;

**Data** task3;

set task2;

title1 '1996 University of Florida Student Survey';

title2 'Agresti and Finley, Statistical Methods for the Social Sciences';

label hsgpa = 'High School GPA'

collgpa = 'College GPA'

polaffl = 'Political Affiliation' ;

**run**;

/\* PROC print is another way to look at a data file besides Proc FSVIEW. \*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Task #4: Print the data file. ;

\* Note: The options command is a system command. ;

\* It is not a part of a Data Step or a Proc. ;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Options to Prevent Date and Page Number Printing ;

Options nodate nonumber ;

\* Options to Set Line Length and Page Size ;

Options ls=**80** pagesize=**54** ;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**Proc** **Print** data=task3 noobs;

\* The var command enables you to list just the variables you want printed ;

var Subject Gender Age HSGPA CollGPA PolAffl Ideology;

**run**;

R Code - Enter the Data

subject <- c(1,2,3,4,5)

gender <- c("m","f","f","f","m")

age <- c(32,23,27,35,23)

HSGPA <- c(2.2,2.1,3.3,3.5,3.1)

CollGPA <- c(3.5,3.5,3.0,3.2,3.5)

task1 <- data.frame(subject, gender, age, HSGPA, CollGPA)

getwd()

uflData <- read.table(file="/Correct Pathname/uflsurvey1996.txt", header=TRUE)

uflData[1,]

uflData[1:5,]

uflData[,1]

uflData$ag

sapply(uflData,mean)