PROJECT 1: ILLUSTRATION OF THE CENTRAL LIMIT THEOREM

/\* Illustrate the CLT\*/

/\* (a) Generate n Uniform[0,1] random variables. \*/

/\* (b) Iterate it "iter" number of times. \*/

/\* (c) For every sample, compute the sample mean. \*/

/\* (d) Plot a histogram for those means. Check that the histogram looks bell-shaped, centered about the mean=0.5, and the inflexion points are 0.5 +/- sqrt(1/12)/sqrt(n).\*/

**%macro** CLT (n, iter);

%do j=**1** %to &iter;

data uniforms;

do i=**1** to &n;

sum+rand("Uniform");

end;

xbar=sum/&n;

keep xbar;

run;

proc append base=allxbars data=uniforms;

run;

%end;

proc univariate data=allxbars;

var xbar;

histogram/normal;

run;

**%mend**;

%***CLT***(**100**, **100**);

PROJECT 2: SALES

/\* (a) Import salesdata.csv. \*/

**proc** **import** out=salesdata

datafile="C:/Users/000110888/Desktop/salesdata.csv"

dbms=csv replace;

**run**;

**proc** **print**;

format date date9.;

**run**;

/\* (b) Plot salesamount against date. \*/

**proc** **gplot**;

plot salesamount\*date;

**run**;

/\* (c) Compute mean salesamount by date. \*/

**proc** **sql**;

create table means as

select date format=date9., mean(salesamount) as msalesamount

from salesdata

group by date;

select \* from means;

**quit**;

/\* (d) Plot mean salesamount by date. \*/

symbol interpol=join value=diamond c=gold width=**3**;

**proc** **gplot** data=means;

plot msalesamount\*date;

**run**;

/\* (e) Compute the overall mean of sale amounts and make

it a macro variable. \*/

**proc** **means** data=salesdata noprint;

var salesamount;

output out=outmeans mean=overallmean;

**run**;

**data** \_null\_;

set outmeans;

call symput('overallmean', round(overallmean, **0.01**));

**run**;

%put &overallmean;

/\* (f) Plot deviations of mean salesamount per date

from the overall mean. \*/

**proc** **sql** number;

create table sales as

select date format=date9., msalesamount-&overallmean as devation

from means;

select \* from sales;

**quit**;

symbol interpol=join value=circle;

**proc** **gplot** data=sales;

plot devation\*date/vref=**0**;

**run**;

PROJECT 3: SALARIES IN A DEPARTMENT STORE

**data** store;

input dept $ name $ salary @@;

datalines;

bedding Watlee 18000 bedding Ives 16000

bedding Parker 9000 bedding George 8000

bedding Joiner 8000 carpet Keller 20000

carpet Ray 12000 carpet Jones 9000

gifts Johnston 8000 gifts Matthew 19000

kitchen White 8000 kitchen Banks 14000

kitchen Marks 9000 kitchen Cannon 15000

tv Jones 9000 tv Smith 8000

tv Rogers 15000 tv Morse 16000

;

/\* (a) Compute the maximum of all the salaries. \*/

/\* (b) Make this maximum a macro variable. \*/

/\* (c) Compute percent of the maximum for each value

of salary. \*/

/\* Alternative route: \*/

**proc** **sql**;

select dept, name, salary, salary/max(salary)\***100** as perc\_max

from store;

**quit**;

/\* With macro variable: \*/

**proc** **means** data=store noprint;

var salary;

output out=result max=maxsalary;

**run**;

**proc** **print** data=result;

**run**;

**data** \_null\_;

set result;

call symput ('maxsalary', maxsalary);

**run**;

%put &maxsalary;

**data** store;

set store;

percsalary=salary/&maxsalary\***100**;

**run**;

title "Maximum salary is &maxsalary .";

**proc** **print** data=store;

**run**;

PROJECT 4: PAIN SCORES

**data** painscores;

input id mos3 mos6 mos12 mos24 mos36;

cards;

1 46 57 67 68 75

2 24 35 . . .

3 30 45 58 . .

4 22 . 48 66 76

5 18 15 14 . .

6 35 67 . . .

7 27 47 50 55 67

8 12 31 41 47 52

9 55 76 . . .

10 18 39 50 67 78

11 10 25 28 33 33

12 35 44 67 . .

13 22 45 . 68 78

14 35 55 70 75 83

;

/\* (a) Make a long-form data set. \*/

**data** longform;

set painscores;

array x[**5**] mos3 mos6 mos12 mos24 mos36;

do visit=**1** to **5**;

score=x[visit];

output;

end;

keep id visit score;

**run**;

**proc** **print** data=longform;

**run**;

/\* (b) Remove all missing values. \*/

**data** longform;

set longform;

if score ne **.** ;

**run**;

**proc** **print** data=longform;

**run**;

/\* (c) Compute the distribution of the number of visits. \*/

**proc** **sql**;

create table nvisits as

select id, count(\*) as n\_visits

from longform

group by id;

select \* from nvisits;

**quit**;

**proc** **sql**;

select n\_visits, count(\*) as npatients

from nvisits

group by n\_visits;

**quit**;

/\* (d) Compute the average number of visits. \*/

**proc** **sql**;

create table meanvisits as

select mean(n\_visits) as mean\_visits

from nvisits;

select \* from meanvisits;

**quit**;

/\* (e) Make the average number of visits a macro variable. \*/

**data** \_null\_;

set meanvisits;

call symput('mean\_visits', round(mean\_visits, **0.001**));

**run**;

%put &mean\_visits;

/\* (f) Print the data with the title "The average number of visits

is <insert the macro variable>" \*/

title "The average number of visits is &mean\_visits .";

**proc** **print** data=painscores;

**run**;