\*Homework 2;

\*question 1;

**PROC** **FORMAT**;

VALUE fmtTreatment

**0** = "Placebo"

**1** = "Progabide"

;

**DATA** Homework2;

INFILE "D:\DATS7510\SAS\Homework2\seizure.txt";

INPUT PID **1**-**2** Seizure1 **9**-**11** Seizure2 **17**-**18** Seizure3 **25**-**26** Seizure4 **33**-**34** Trt **41** BaseSeiz **49**-**51** Age **57**-**58**

;

LABEL PID = "Patient Identifier"

Seizure1 = "Seizure Count Week 1"

Seizure2 = "Seizure Count Week 2"

Seizure3 = "Seizure Count Week 3"

Seizure4 = "Seizure Count Week 4"

Trt = "Treatment"

BaseSeiz = "Baseline Seizure Rate"

Age = "Age of the Patient";

FORMAT Trt fmtTreatment.;

\*put keyword LABEL in PROC PRINT line to output labels;

**PROC** **PRINT** DATA=Homework2 LABEL;

\*SUM BaseSeiz;

**RUN**;

\* (a) Using PROC FREQ determine how many patients were randomized to the placebo and

how many were randomized to the progabide treatment. Answer after your output.;

**PROC** **FREQ** DATA=Homework2;

TABLE Trt;

**RUN**;

\* (b) Using PROC MEANS to answer - what are the N, MEAN, and SD of seizure rate at baseline?;

**PROC** **MEANS** DATA=Homework2 N MEAN STDDEV;

VAR BaseSeiz;

**RUN**;

\* (c) Using PROC MEANS answer, what are the mean, standard deviation and median seizure rates within

each treatment group at each week following treatment? Answer after your output.;

**PROC** **SORT** DATA=Homework2;

BY Trt;

**PROC** **MEANS** DATA=Homework2 MEAN STDDEV MEDIAN;

\*CLASS Trt;

VAR Seizure1 Seizure2 Seizure3 Seizure4;

BY Trt;

**RUN**;

\* (d) Using PROC PLOT, produce a plot of the seizure rate at week 4 (Seizure 4) vs the baseline seizure rate (BaseSeizure)

for each treatment group.;

**PROC** **SORT** DATA=Homework2;

BY Trt;

**PROC** **PLOT** DATA=Homework2;

PLOT Seizure4\*BaseSeiz="O";

BY Trt;

**RUN**;

\*question 2;

\*The variable names are in the parentheses, create a LIBNAME statement to set up a library where this data is stored.

Create a temporary dataset in your program and use the SET statement to set the permanent SAS data set. Please

label each variable name and format them.;

\*a) LABEL each variable, and run a PROC CONTENTS, to check each variable;

**PROC** **FORMAT**;

VALUE fmtfos

**1** = "yes"

**2** = "no"

;

VALUE fmtswimlocation

**1** = "non-beach"

**4** = "beach"

;

VALUE fmtagegrp

**2** = '15-19'

**3** = '20-24'

**4** = '25-29'

;

VALUE fmtsex

**1** = "male"

**2** = "female"

;

**RUN**;

LIBNAME swim "D:\DATS7510\SAS\Homework2";

**DATA** swimdata;

SET swim.swimmers;

FORMAT fos fmtfos. swimlocation fmtswimlocation. agegrp fmtagegrp. sex fmtsex.;

LABEL fos = "Frequent Ocean Swimmer"

swimlocation = "Swimming Location"

agegrp = "Age Group"

sex = "Sex"

numearinfect = "Number of Ear Infections"

;

\*put keyword LABEL in PROC PRINT line to output labels;

**PROC** **PRINT** DATA=swimdata LABEL;

\*SUM numearinfect;

**RUN**;

\*a) LABEL each variable, and run a PROC CONTENTS, to check each variable.;

**PROC** **CONTENTS** DATA=swimdata;

**RUN**;

\*b) Use PROC FREQ to determine the frequency distribution of each categorical variable overall

(\*some variables are numeric form about are categorized should also count frequency). Use PROC MEANS

to determine the mean, standard deviation and median number of ear infections overall. Add titles in your output.;

**PROC** **FREQ** DATA=swimdata;

TABLES fos swimlocation agegrp sex;

**RUN**;

\*Use PROC MEANS to determine the mean, standard deviation and median number of ear infections overall.

Add titles in your output.;

**PROC** **MEANS** DATA=swimdata MEAN STDDEV MEDIAN;

VAR numearinfect;

TITLE "Mean, Stddev, Median, and Table of swimmers";

**RUN**;

\*c) Use one PROC FREQ to produce cross tabulations of swimming location by age group and by sex group.

Among each swimming location, which age group is the highest proportion? Among each swimming location, which

sex group is the highest proportion?;

\*PROC SORT DATA=swimdata;

\*BY swimlocation;

**PROC** **FREQ** DATA=swimdata;

TABLES swimlocation\*(sex agegrp);

\*BY swimlocation;

\*WHERE swimlocation=1;

\*WHERE swimlocation=4;

**RUN**;

\*d) Use PROC MEANS to produce the mean, stddev and median number of ear infections in each swimming location

group, OUTPUT the results into a new temporary data set named "averageearinf".;

**PROC** **SORT** DATA=swimdata;

BY swimlocation;

**PROC** **MEANS** DATA=swimdata MEAN STDDEV MEDIAN;

\*CLASS swimlocation;

VAR numearinfect;

BY swimlocation;

OUTPUT out=averageearinf;

**RUN**;