Homework 2

Write the code for each of the problems and upload the assignment on Sakai. You should include comments with your code: name, date, explanation of the code at the top of each problem and any other places comments may be necessary. Be sure to include titles on output.

1. Adapted Problem 2.2 from the text

a) Using the following data, create a SAS data set called PATIENT using the column input method. The column layout is below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Description** | **Starting Column** | **Length** | **Type** |
| ID | Subject ID | 1 | 3 | Char |
| GENDER | Gender | 4 | 1 | Char |
| RACE | Race of subject | 5 | 1 | Char |
| AGE | Age of the subject | 6 | 2 | Num |

DATA:

001MW35

002FW41

003MB62

004FB38

005MW44

006FB47

007FW53

008MW58

009FB56

010FB39

b) Using the second set of data, create a SAS data set called PATIENTVITALS.

In the DATA step, include a statement to compute the average blood pressure (call it AVE\_BP) computed as the diastolic blood pressure (DBP) plus one-third the difference of the systolic blood pressure (SBP) and the DBP (For those who are interested in this, since the heart spends more time in diastole, relaxed state, than it does in systole, when it contracts, a weighted average is used to represent the “average” blood pressure – one-third of the SBP and two-thirds of the DBP).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Description** | **Starting Column** | **Length** | **Type** |
| ID | Subject ID | 1 | 3 | Char |
| HR | Heart rate | 4 | 3 | Num |
| SBP | Systolic blood pressure | 7 | 3 | Num |
| DBP | Diastolic blood pressure | 10 | 3 | Num |
| N\_PROC | Number of procedures | 13 | 2 | Num |

DATA:

00108013008010

00208811007205

00305018810002

004 10806801

00506812208204

006101 07404

00707810406603

00804811207006

00907719011009

01006616410610

c) Use PROC MEANS to compute the number of nonmissing values, the mean, the standard deviation, the 95% CI on the mean, and the median value for SBP, DBP, and AVE\_BP. The data layout:

d) Merge the two data sets PATIENT and PATIENTVITALS. Create a side-by-side box plot of the heart rate (HR) by race.

**Problem 1 grading:**

**2 points for entering data with column input style.**

**2 points for calculating the values from Proc Means**

**1 point for merging data**

**3 points of side-by-side box plot**

2. Use SAS to open the file weather.xls which contains 11 years of weather data (note there are missing dates) forChicago courtesy of http://www.wunderground.com. -you will need to review the Proc Import procedure for this exercise. Be sure you open the file in Excel first to see what is in the file and the name of the worksheet.

1. Find the average temperature, average high and average low of Chicago. Include the standard deviation, variance and 95% CI.

Why do you think the standard deviation (or variance) is so big for all three and in

particular for the average high temperature?

**OUTPUT: TABLE WITH AVG, ST DEV, VAR and 95% CI.for each temp**

1. Find the average winter temperature of Chicago (let winter be defined as December, January and February).

**OUTPUT: TABLE WITH AVG, ST DEVand 95% CI.for winter**

Why is the standard deviation smaller on the winter months than the entire year?

1. Create three histograms, one for the mean wind speed, max wind speed and max gust. (Turn in the code and the histograms, not the other data). Include the normal curve on the graph. Inset the mean, standard deviation, skew and kurtosis to 3 decimal places.  
   Describe the skew and kurtosis of the three histograms.

**OUTPUT: HISTOGRAM with normal curve and inset, Normality Test Table**

**Problem 2 Grading**

**2 points for properly opening excel file and creating data set**

**2 points for part a) (don’t forget the analysis question)**

**2 points for part b.**

**6 points for creating and analyzing the histograms**

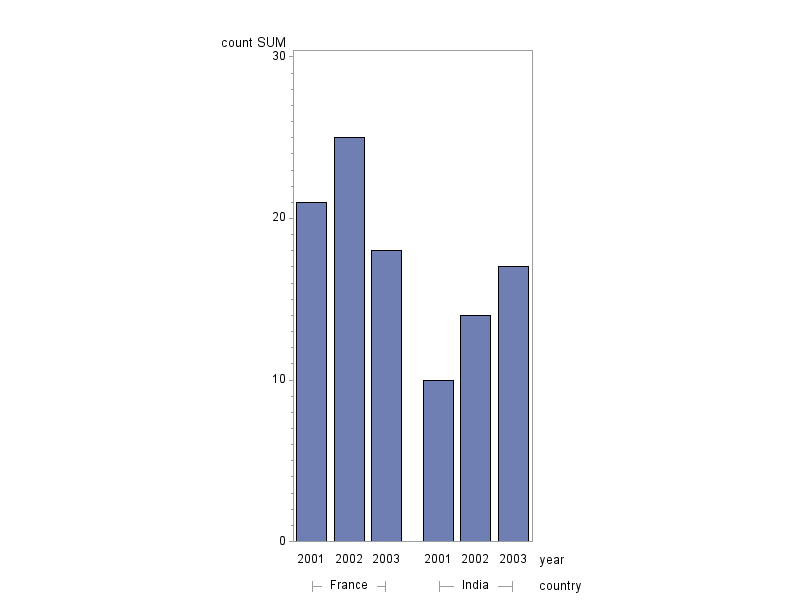
3. The file adoption.txt contains data from 2003 through 2013 on total international adoption. (source<http://adoption.state.gov/about_us/statistics.php>) Write the data code to open and upload the data for country, year and adoptions. Output: graphs asked for and code.

a) Create a horizontal bar chart for the total number of adoptions in each year.. (Make sure your vertical axis (independent variable) contains the years being considered, not some other bin sizing). If you include

length country $ 12 ;

before the input statement in the data step (and you name your variable country), you will get the full name of each country.

b) Create a side-by-side vertical bar chart for the total number of adoptions for each year separated by country. I want 2003 – 2010 bars for China (or whatever country) right next to each other, then 2003 – 2010 for the next country as shown below. ( Again, make sure x-axis contains all 6 years) .



c) Create one pie chart for the total adoptions over all 6 years with the pie slices being the total number of adoptions from country.

d) Create a scatter plot of the number of adoptions over each year from only these countries: China, Russia, Guatemala and Ethiopia. To create a trend line you will include the command I = Join at the end of the symbol value call. (This stands for interpolation = join).  
Ex: symbol1 value = "circle" color = blue I = join;

e) Answer the following questions. You may need to do a little research on this.

1). What is the overall trend in international adoption over the past 8 years? Which graph(s) support this? Why do you think this occurring?

2) Which country has seen an increase in adoptions over the years? Which graph(s) support this? What might be precipitating this increase?

3) Which countries have seeninternational adoption go to almost none or none? Which graph(s) support this? Why has this happened? (Very disturbing)

**Problem 3 Grading**

**2 pts for data step**

**Part a) 4pts**

**Part b) 4 pts**

**Part c) 3 pts**

**Part d) 4 pts**

**Part e) 3pts- 1 pt for each question**