Lab 5 - Choropleth Mapping **Name** _____

In this exercise, you will prepare a choropleth map illustrating rates for Alcohol Related Fatalities for the United States. The data source is the same as the last lab. To do this exercise, calculate the rate (cases per 100,000 population). Then calculate class limits for 4 classes according to the following class limit schemes:

1. Natural Breaks

Determine class intervals using a natural breaks method only after examining the data thoroughly. Rank the data in descending order, and scan them for the largest arithmetic breaks between adjacent values. These breaks will serve as class delimiters. The values you should use for your class limits will be the value halfway between two adjacent values that are in different classes.

The class limits are:

_____ — _____

_____ — _____

_____ — _____

_____ — _____

2. Quantiles

By dividing the data into quantiles (specifically quartiles for this exercise), you force the same number of observations to fall within each class (excluding remainders). To divide the data into quartiles:

a. rank the data, from high to low;

b. calculate the upper limits using the following equation:

$UCL_n = nk$ where:

UCL_n is the upper limit of the nth class
n = the number of the class for which the upper limit is desired
k = the number of observations/number of classes;

c. using the value calculated for UCL_n, count up from the low end of the ranking and mark off the first class boundary immediately after the UCL_nth observation. Continue counting from the class boundaries to mark of subsequent class boundaries;

d. the class intervals are:

_____ — _____

_____ — _____

_____ — _____

_____ — _____

3. Equal steps based on the data range

Using this method, you will force the same ranges of data values to fall within each class, yet you cannot guarantee that each class will hold the same number of observations. To determine the class limits:

a. calculate the range of the data, R, where:

R = H - L where

H equals the highest observation and L equals the lowest observation;

b. find the common difference (spread) for each class, CD, where:

CD = R/number of classes;

c. determine the upper class limits, UCL_n where:

$UCL_n = L + n(CD)$, n = class number;

d. the class limits are:

_____ — _____

_____ — _____

_____ — _____

_____ — _____

4. Measures of central tendency

You will now use the mean and standard deviation of the data to determining class limits.

From statistics, we know that the:

mean = $\sum X/n$, or the sum of the observations divided by the number of observations.

standard deviation (SD) = $\sqrt{\sum(\text{mean-X})^2/n}$, or the square-root of the sum of the (mean minus each observation squared), divided by the number of observations.

4.1 Standard deviation

a. calculate the mean and standard deviation for the data. Form class limits as follows:

class 1: < -1SD
class 2: -1SD to mean
class 3: mean to +1SD
class 4: > +1SD

b. if these data are normally distributed, then one would expect classes 2 and 3 to each contain 34% of the observations and classes 1 and 4 to each contain 16% of the observations. For the US, then, we may test the normalcy of the data by predicting how many observations should fall within each class. Classes 2 and 3 should each contain 17 (rounding down) states, and classes 1 and 4 should each contain 8 (rounding down) states. ***How does the distribution of the Alcohol Fatalities data compare***? (answer on next page)

c. the class intervals for this method are:

_____	—	_____
_____	—	_____
_____	—	_____
_____	—	_____

4.2 Nest Means

We will now use the method of nested means to classify the data.

a. to find the class divisions, first take the mean of the entire data set. Using the mean as the first subdivision of the data, take the means of the upper and lower halves of the data. Use the values for the means as class limits;

b. the resulting class limits are:

_____	—	_____
_____	—	_____
_____	—	_____
_____	—	_____

5 Jenk's Optimization Method

The goal of using this method is to form classes that are internally homogenous while assuring heterogeneity among classes. We will use Jenk's method to analyze the appropriateness of the 5 classifications you have just performed. Here is the procedure:

1. Compute the mean of the entire dataset, and calculate the sum of the squared deviations of each observation:

$$\sum(\text{mean-X})^2$$

This amount is called the SDAM (squared deviations, array mean);

2. Using the class boundaries calculated in parts **1-4** for this exercise, compute the class means (this means computing 4*8 class means);

3. Calculate the deviations of each X in each class from its class mean (mean-X), and square these;

4. Sum the squared differences in the previous step for each classification (this will be called the SDCM (squared deviations, class means).

5. Compute goodness of variance fit (GVF)

$$GVF = \frac{SDAM-SDCM}{SDAM}$$

GVF natural breaks _____

GVF quantiles _____

GVF equal steps _____

GVF standard deviation _____

GVF nested means _____

6. Values should range from 0.0 to 1.0. You are looking for the value closest to 1.0 as the best fit.

7. Which classification scheme calculated in parts **1-4** has the best goodness of fit? Do you think this scheme best represents the data? Why? (answer this in the one page written justification indicated below).

What you should turn in:

This handout, completed;

Select the classification scheme above that you feel best portrays Alcohol Fatalities rates by state. The purpose is to construct an honest and informative portrayal of the distribution of the data. *Write a one page justification for your choice*;

Your compilation worksheet (optional);

Construct a choropleth map based on the classification scheme you feel best portrays the data, adhering to cartographic principles covered in previous exercises, and discussed in class and in the text. Print a copy of this map to include with the above items.