

## Chapter 13 Project: Taxes on Toxics

Name \_\_\_\_\_ Name \_\_\_\_\_

### Overview

In this project you will investigate the amount of RCRA waste produced by several counties in a state. RCRA waste is solid waste assigned a federal hazardous waste code and is regulated by RCRA—the Resource Conservation and Recovery Act. (Note: RCRA is pronounced “wreck-ruh”.) You will devise a strategy to fine counties that produce lots of hazardous waste, and reward counties that produce lesser amounts of hazardous waste. The concepts of normal distributions and z-scores will play an integral role in your strategy.

### Obtain the Data

Your instructor will provide you with a printed data sheet listing RCRA and population data for 1 of 12 states. Your state: \_\_\_\_\_

In StatCrunch, enter manually the Mean RCRA Waste data into var1, and the population data into var2. Re-label the column headers to “tons” and “people.” Alternately, download your state data directly from [Enviromath.com](http://Enviromath.com).

### 1. Analyze the RCRA data

a) Listed on the data sheet are the counties in your state that produce the greatest amount of RCRA waste. Production values are given for the years 1991, 1993, 1995 and 1997. For the first 2 counties on the data sheet, compute the mean RCRA waste amount from 1991 to 1997. Show work below.

b) Do your values agree with those listed in the column “Mean RCRA Waste”?  
\_\_\_\_\_ What are the units of measure for the mean?

c) Scan the data to find a county with a *huge* change in RCRA waste generation from one biennium to the next: which county, how much waste one year, how much in the next year? What is the percent change from one biennium to the next?

d) Such extreme changes in hazardous waste production do not seem reasonable, maybe the numbers are in error. But maybe not! Give one *reasonable explanation* why RCRA waste generation might change so much in one biennium.

e) For the mean RCRA waste values in column 1, compute the following statistics. Round to the nearest ton. *Recall that these values can be found by selecting **Stat > Summary Stats > Columns**. Then select **tons** and click on **Calculate**.*

sample size  $n =$  \_\_\_\_\_

minimum = \_\_\_\_\_ maximum = \_\_\_\_\_

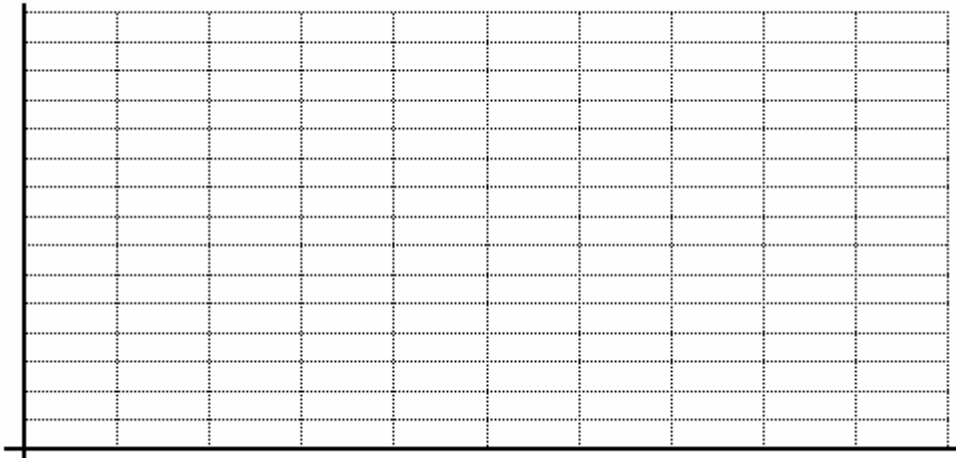
mean  $\bar{x} =$  \_\_\_\_\_ standard deviation  $s =$  \_\_\_\_\_

The *total* of all values in a column can be found by multiplying the mean ( $\bar{x}$ ) by the number of data values ( $n$ ). What is the total of the mean RCRA waste produced by the counties on your data sheet?

f) Make a frequency histogram of the mean RCRA waste values in column 1. Your histogram should have 5-10 bins. *Below are the steps to create a histogram on StatCrunch.*

1. Select **Graphics > Histogram** and then select the column **tons**.
2. Select **Next** to set up bins manually and control other options, or select **Create Graph!** to have StatCrunch do the work for you.
3. To modify the histogram, click **Options > Edit** and change the binning parameters, labels, etc. Press **Create Graph!** to return to the histogram.

Sketch the histogram below, or attach a printout; label and scale clearly.



g) Inspect the histogram. Are the mean RCRA waste values normally distributed? Explain

h) For the mean RCRA waste data, compute the 7 numbers:  $\bar{x} - 3s$ ,  $\bar{x} - 2s$ ,  $\bar{x} - s$ ,  $\bar{x}$ ,  $\bar{x} + s$ ,  $\bar{x} + 2s$ , and  $\bar{x} + 3s$ . Round to the nearest ton.

i) Do any of the 7 numbers come out negative? If so, do these numbers have any physical meaning? Can you have negative mean RCRA waste in reality?

j) Sometimes there are data that seem to be “way out of bounds.” Statisticians call these numbers **outliers**; outliers are numbers that lie at least 3 standard deviations away from the mean. Are there any outliers in your mean RCRA waste values? If so, what are the names of the counties?

You probably noticed how outliers tend to dominate calculations and skew histograms. Sometimes we know that outliers are caused by error, and so we can delete them from the data set. But sometimes they are accurate, and we must leave them in. In the RCRA data, you don't know whether your outliers are accurate or erroneous, so just leave them in! For the bonus question of this project (on the last page) you can investigate the source of the largest outlier.

## **2. Per Capita Waste**

a) The EPA hires you as a consultant, to impose fines on counties that are “environmentally bad.” Your supervisor suggests that counties that *generate the most RCRA waste* should be fined the most. Discuss why this system might not be fair.

Another approach is to fine the counties that produce a lot of waste relative to their population size. In other words, fine the counties that have the highest mean RCRA waste **per capita** (per person).

b) For the first 2 counties listed on your data sheet, compute (by hand) the mean RCRA waste per capita. Convert the result so that the units are in **pounds per person**. Show work below.

You would like to repeat the above calculation for all counties on your data sheet. To have StatCrunch do the “heavy lifting,” select **Data > Transform data**. In the transformation box at the top of the window, type in **2000\*tons/people** and then press **Transform**.

The per capita RCRA values will be computed and stored in column 3, with all values having units of “pounds per person.” Re-label the header on the new column to **lbs/person**. Then select **Data > Save data**.

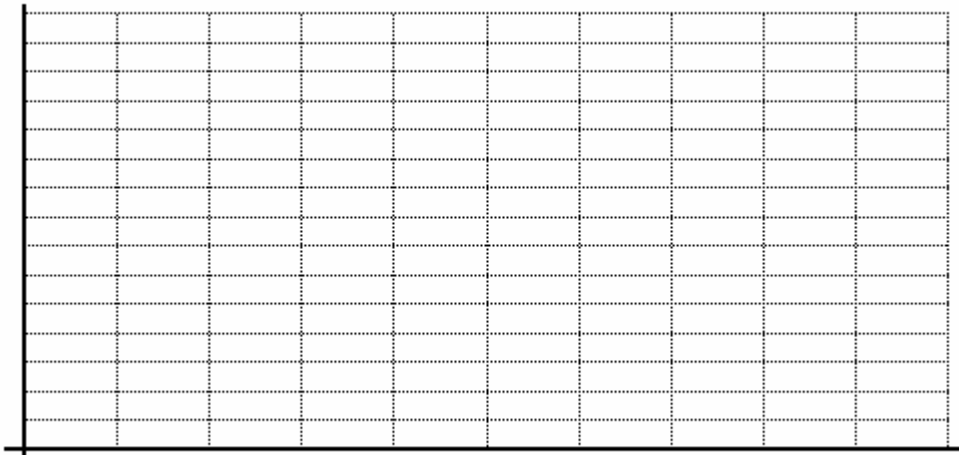
c) Do the first 2 entries in column 3 on StatCrunch match what you computed earlier? \_\_\_\_\_ Record the column 3 values on your data sheet using 1 decimal place of precision.

d) For the per capita mean RCRA values in column 3, compute the following statistics. Round to 1 decimal place.

minimum = \_\_\_\_\_ maximum = \_\_\_\_\_

mean  $\bar{x}$  = \_\_\_\_\_ standard deviation  $s$  = \_\_\_\_\_

e) Make a frequency histogram of the per capita mean RCRA waste values. Again use 5-10 bins. Sketch the histogram below, or attach a printout. Label and scale axes appropriately.



### 3. Carrots and Sticks

You have normalized the county data by computing the per capita RCRA waste. Now you come up with the idea to penalize the heavy-polluting counties with **waste fines** and reward the light-polluting counties with **waste credits**. Here's how your plan will work.

Counties that have RCRA waste production above the statewide mean will be fined, and those with RCRA production below the statewide mean will be given waste credits. To reward and penalize the counties on a **continuous** scale, you decide to base the fines and credits on z-scores. Recall that a z-score will indicate how many standard deviations a county lies above or below the statewide mean. Z-scores are computed with the simple formula:

$$z = \frac{x - \bar{x}}{s}$$

In the formula,  $x$  is each county's per capita mean RCRA waste,  $\bar{x}$  is the statewide mean of per capita waste values, and  $s$  is the standard deviation of the statewide per capita waste values.

a) Compute the z-score of the per capita waste values for the first 2 counties on your data sheet. Show work below.

You can have StatCrunch determine z-scores for all per capita county values. First have StatCrunch compute (again) the mean and standard deviation of column 3. Record values here using *all decimal places*.

$$\bar{x} = \underline{\hspace{4cm}} \quad s = \underline{\hspace{4cm}}$$

Now select **Data > Transform data**. In the transformation box at the top of the window, type in ("**lbs/person**" -  $\bar{x}$ ) /  $s$  using all decimal places for  $\bar{x}$  and  $s$ . (You *must* include the quotes around lbs/person!!) Then press **Transform**.

A 4th column will be added with the z-scores. Change the column header to **z-score**. Then save your work by selecting **Data > Save data**.

- b) Inspect the column of z-scores in StatCrunch. Do the first 2 entries match what you computed earlier? \_\_\_\_\_ Record county z-score values in the first z-score column on the data sheet; round values to 2 decimal places.
- c) Your boss thinks your “carrots and sticks” strategy based on per capita z-scores has promise, but she is worried that there might be an imbalance between the number of counties receiving credits and the number getting fined. Is she right? Explain.

#### **4. Transformation to Normal**

Your boss firmly believes that the number of fines and the number of credits should be approximately equal. You know that this is impossible because the per capita waste values for the counties on your data sheet have a large positive skew.

You consult your text and find that positively skewed data can often be made more symmetrical by taking the logarithm of the data values. You decide to revise your data using that strategy.

- a) Compute the logarithms of the per capita mean RCRA waste values for the first 2 counties on your data sheet. Show work below.

Again you want to have StatCrunch do the rest of the work. To compute the logarithm of the mean per capita RCRA values and store the results in column 5, select **Data > Transform** data. In the transformation box at the top of the window, type in **log10("lbs/person")**. Again, be sure to include quotes around lbs/person! Then press **Transform**. Finally, save your work by selecting **Data > Save data**.

- b) Inspect the column of logged values. Do the first 2 entries match what you computed earlier? \_\_\_\_\_ Record the logged values on your data sheet using 2 decimal places of precision.

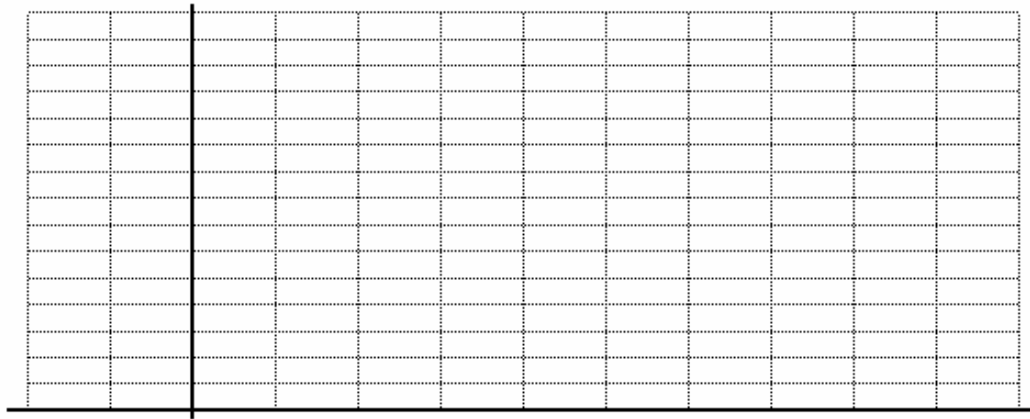
c) For the logged per capita mean RCRA values, compute the following statistics. Round to 2 decimal places.

minimum = \_\_\_\_\_ maximum = \_\_\_\_\_

mean  $\bar{x}$  = \_\_\_\_\_ standard deviation  $s$  = \_\_\_\_\_

d) What are the units of measure for these statistics?

e) Make a frequency histogram of the logged per capita values using 5-10 bins. Sketch below or attach a printout; include units and labels. *Note: some of the logged values may be negative!*



f) How does the histogram of the transformed (logged) data compare to the two histograms that you have sketched previously? Explain.

g) For the transformed (i.e. logged) values in column 5, calculate the 7 numbers:  $\bar{x} - 3s$ ,  $\bar{x} - 2s$ ,  $\bar{x} - s$ ,  $\bar{x}$ ,  $\bar{x} + s$ ,  $\bar{x} + 2s$ , and  $\bar{x} + 3s$ . Round to 2 decimal places.



h) Use these 7 numbers and the Empirical Rule to help you determine if the transformed data are approximately normal. Show work.

### **5. Carrots and Sticks Revised**

You have transformed the per capita county waste data into a distribution that is closer to normal, and certainly more symmetrical. You would now like to return to the idea of computing z-scores for each county to determine how much each will be penalized or rewarded.

a) Compute the z-scores for the logged per capita values for the first 2 counties in column 5. Show work below.

b) Compute the z-score for the logged per capita values (in column 5) and store the results in column 6. Use StatCrunch for assistance as you did before. In part of the transformation you'll need to type **log10("lbs/person")**. Be sure to save your work when finished.

When finished, re-label the new column with **z-score**. check that the first 2 entries match up with your earlier computations. Do they? \_\_\_\_\_ Copy the z-scores from column 6 to your data sheet using 2 decimal places of precision.

c) Inspect your column of logged per capita z-scores. How many counties will be given fines, and how many will be given credits?

d) You're feel pretty good about this revised carrots and sticks system, as you will be rewarding about the same number of counties that you are penalizing. You show it to your boss who thinks it's great too. She now gives you enough money to impose fines and give credits. She suggests a \$100,000 fine or credit per z-score (fines for positive z-scores, credits for negative z-scores). Will your agency lose money, earn money, or break even? Explain in detail.

## **6. BONUS**

Earlier in this project you listed the counties in your state that produce an inordinate amount of RCRA waste. Which county is the most extreme outlier? \_\_\_\_\_ In which biennium did that county produce the most RCRA waste?\_\_\_\_\_

Detailed information about RCRA waste production is provide at the Right-to-Know Network: <http://www.rtknet.org/>. Go to this website, click on **DataBases** and then select **Biennial Reporting System (BRS)**. In the BRS database, you will search by **geographic area**. In the BRS search window, select the state and the county and year that you listed above, then press **Submit Query**.

The database will create a long table of companies in your outlier county that produce RCRA and other hazardous waste. Scroll through the table and find a company that **generated** one of the greatest amounts of waste.

Click on the company name, and you will get another screen of information. Search through the list of waste products and find the type of waste that was produced in the greatest amount. (Each table will list "waste description" and "tons generated current year.")

Summarize your findings: What is the name of the company that was producing the greatest amount of waste in your outlier county? How much waste did the company produce? What type of waste did the company produce in the greatest amount? What was the amount produced?