Synthesis through Service Learning in Statistics
Part One

Gina Reed
Gainesville College

In 1992, efforts towards undergraduate statistics education reform began to emerge. The Mathematical Association of America (MAA) published *Statistics for the Twenty-First Century, Perspectives on Contemporary Statistics,* and *Heeding the Call for Change*. In these volumes, many statisticians agreed there was a gap between statistics teaching and statistical practice. The American Statistical Association/ Mathematical Association of America Joint Curriculum Committee made three recommendations to improve the teaching of statistics:

- emphasize the elements of statistical thinking
- incorporate more data and concepts
- foster active learning

One of the most exciting possibilities that incorporates the recommendations of the American Statistical Association/ Mathematical Association of America Joint Curriculum Committee is service-learning. Service-learning is a concrete application of statistical methods using real data with the analysis and interpretation that is useful to a community agency. Service-learning projects emphasize the elements of statistical thinking because using real data gives students the opportunity to explore the statistical concepts presented in class and receive practical experience at the same time.

Heights of Presidential Candidates

How would you compare the heights of the two groups?

Year	Winner	Height (inches)	Runner-Up	Height (inches)
1900	McKinley	67	Bryan	72
1904	T. Roosevelt	70	Parker	72
1908	Taft	72	Bryan	72
1912	Wilson	71	T. Roosevelt	70
1916	Wilson	71	Hughes	71
1920	Harding	72	Cox	*
1924	Coolidge	70	Davis	72
1928	Hoover	71	Smith	*
1932	F. Roosevelt	74	Hoover	71
1936	F. Roosevelt	74	Landon	68
1940	F. Roosevelt	74	Wilkie	73
1944	F. Roosevelt	74	Dewey	68
1948	Truman	69	Dewey	68
1952	Eisenhower	70.5	Stevenson	70
1956	Eisenhower	70.5	Stevenson	70
1960	Kennedy	72	Nixon	71.5
1964	Johnson	75	Goldwater	72
1968	Nixon	71.5	Humphrey	71
1972	Nixon	71.5	McGovern	73
1976	Carter	69.5	Ford	72
1980	Reagan	73	Carter	69.5
1984	Reagan	73	Mondale	70
1988	Bush	74	Dukakis	68
1992	Clinton	74	Bush	74
1996	Clinton	74	Dole	74
2000	Bush	71	Gore	73

Source: Paul M. Sommers, "Presidential Candidates Who Measure Up," *Chance* 9, no. 3 (1996): 29–31.

1. Construct dotplots of the two groups. Can you determine which group is taller?

2. Find the mean and median heights of the two groups.

3. Based on comparing the mean versus the median, describe the shapes of the two groups.

4. Would the standard deviation or the IQR be the best measure of the variability (based on the shape)? Find the appropriate measure for both groups.

5. Using the above information, write a paragraph comparing the heights of the two groups.

Lab 1

Type all answers in MiniTab session, and turn in lab at the end of class. Remember to type in your name at the top of the session!

1)   Go to http://www.usg.edu/pubs/info_digest/2001/academic/12.phtml, and print the SAT scores. You will need to enter the data in MiniTab. In column 1, enter **1** if the SAT is from a state university and enter **2** if it is from a two-year college. In column 2, enter the SAT scores for the year 2000 (the last column of data on the webpage).

2)   Construct histograms for column two.
                Click on **Graph**.
                Click on **Character Graph.**
                Click on **Boxplot**.
                Enter C2 in variable box.
                Click on box beside **By variable**.
                Enter C1.
                Click on **Ok**.

3)   Describe the shape of the boxplot for state universities.

4)   Describe the shape of the boxplot of two-year colleges.

5)   Compare the similarities and differences of the SAT scores from the two types of institutions based on their boxplots.

6)   Find the descriptive stats for column two.
                Click on **Stat**.
                Click on **Basic Statistics**.
                Click on **Descriptive Statistics**.
                Enter C2 in variable box.
                Click on box beside **By variable**.
                Enter C1.
                Click on **Ok**.

7)   Compare the similarities and differences of the SAT scores from the two types of institutions based on their means and medians.

8)   Compare the similarities and differences of the SAT scores from the two types of institutions based on their standard deviations and IQR (IQR=Q3-Q1)

DATASHEET

Student	Male-0 Female-1	Height (inches)	Hours of sleep	Miles to GC from home	SAT score	Amount of change	Current GPA
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

Using the datasheet, find the average SAT score for your class.

Is it close to the SAT average(966) for Gainesville College on
http://www.usg.edu/sra/students/sat/ftf_sat_fall_2004.pdf?

Perform a hypothesis test to determine if the class SAT average is higher than 966 (i.e. the overall college SAT average).

H_o

H_A

test statistic

p-value

Conclusion

Birthrate and GNP

Country	Birthrate	GNP
Algeria	29.0	1.6
Argentina	19.5	4.0
Australia	14.1	16.6
Brazil	21.2	2.6
Canada	13.7	20.8
China	17.8	1.3
Cuba	14.5	1.6
Denmark	12.4	24.2
Egypt	28.7	0.5
France	13.0	24.1
Germany	11.0	19.8
India	27.8	0.3
Iraq	43.6	0.7
Israel	20.4	13.6
Japan	10.7	27.3
Malaysia	28.0	2.5
Mexico	26.6	3.1
Nigeria	43.3	0.2
Pakistan	41.8	0.4
Philippines	30.4	0.7
Russia	12.6	8.6
South Africa	33.4	2.6
Spain	11.2	13.4
United Kingdom	13.2	17.4
United States	15.2	22.6

Source: *Statistical Abstract of the United States,* 1995.

1. Which variable is the explanatory variable? Explain your answer.

2. Construct a scatterplot.

3. Is the relationship strong or weak? Positive or negative? Explain the relationship in your own words.

4. Can you determine whether the correlation is positive or negative? Explain.

5. Find the correlation and write a sentence about it.

6. Find the coefficient of determination and write a sentence about its' meaning.

7. Find the slope and explain its meaning.

8. Find the regression line and draw it on the scatterplot.

9. Find and circle the point on the scatterplot representing the U.S.

10. Does the regression line underpredict or overpredict the GNP? Will the residual be positive or negative?

11. Predict the GNP for the U.S.

12. Find the residual for the U.S.

Useful websites

http://exploringdata.cqu.edu.au/

http://lib.stat.cmu.edu/DASL/

http://www.amstat.org/publications/jse/jse_data_archive.html

http://www.dartmouth.edu/%7Echance/teaching_aids/data.html

http://www.census.gov/statab/www/ranks.html

http://www.nationmaster.com/

Useful Applets

http://www.stat.uiuc.edu/courses/stat100/cuwu/datalist.html

http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/index.html

http://www.ltcconline.net/greenl/java/Statistics/HypTestProp/HypTestProp.htm

http://statweb.calpoly.edu/chance/applets/Reeses/ReesesPieces.html