A note on Prediction of the U.S. Population and Diabetic Population

B Enye, H. Oh, M. Lemma, and E. McBeth

Department of Mathematics, Savannah State University, USA

The world is home to over seven billion people, and more than twenty million animals. With seven continents, the population of the world is spread out in each country within the continents. One of the most populated countries in the world is the United States. It is ranked 3rd on the list of largest countries by population, and is home to over 300 million people. Figure 1, show the population of the United States from the 1980 to 2014, and those keep increasing every day. Although, the country is for its entertainment, scientific research, and multicultural diversity because people from all over the world move to have a better life, is also has the highest rate of diabetes on the list of developing countries. Figure 2, show the number of diabetics cases from 1980 to 2014. The data is based on age range (18 and higher), education (less than to greater then high school), gender, and race. Combining the different categories together, the data and graphical trend was created. Although the rate, is from 1980 to 2014, the goal of the research is to predict the population of the United States and the number of diabetic cases within that population for 2018 using Least Square Approximation Method.
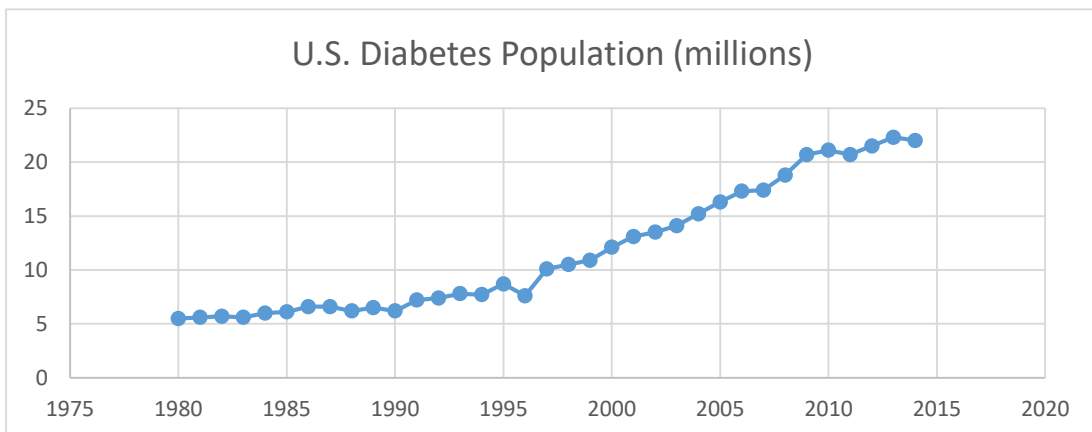

Figure 2: The United States diabetic population from 1980 to 2014. Data is based on age, gender, race, and education
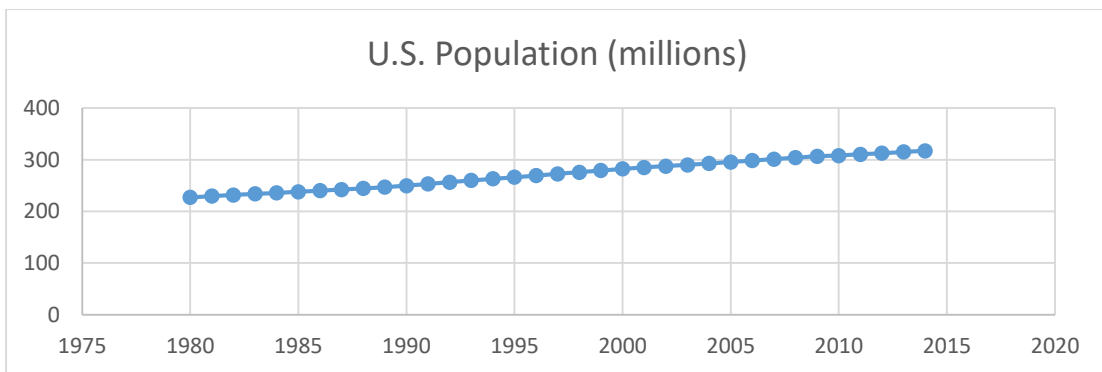

Figure 1: The United States Population from 1980 to 2014. The graph uses date based on age, gender and education

Least Square Approximation Method

This method involves using a standard method in regression analysis to the estimate the exact values/solutions of overdetermined systems. This means that there are calculations which there are more equations than unknowns. The approximation polynomial formula is

$$P_n(x) = \sum_{k=0}^{n} a_1 x^k, with \; n < m - 1$$

The error value formula is

$$E = \sum ((y_i - P_n(x_n))^2$$

Combining the approximation polynomial formula and error value, the Least Square Approximation formula is

$$P_n(x) = an * x^n + an * x^{n-1} + \cdots a_0$$

$$\sum_{i=0} a_n x^k$$

Just like an equation could be linear, quadric, or cubic, least approximation approximate could apply the same concept. The Linear Approximation error

$$\sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$$

Quadric Approximation error

$$\sum_{i=1}^{n} (y_i - (axi^2 + bxi + c))^2$$

Cubic Approximation error

$$\sum_{i=1}^{n} (y_i - (axi^3 + bxi^2 + cxi + d))^2$$

Objective

Using the Least Square Approximation Method, we will be able to predict the increase or decrease in the U.S. general population and the number of diabetic patients for 2018.

Materials and Methods

The data for the general population and the diabetes population were obtained from the census website and Center of Disease Control and Prevention (CDCP) respectively. Table 1, shows the years ranging from 1980 to 2014 for the population needed for the research. Using this table, a mathematical software known as Maple 16 was used to generate the images and graph information in Table 2 and Table 3.

Table 1: U.S General Population and Diabetic population in millions

Year	U.S. Diabetes Population (millions)	U.S. Population (millions)
1980	5.5	227.22
1981	5.6	229.47
1982	5.7	231.66
1983	5.6	233.79
1984	6	235.82
1985	6.1	237.92
1986	6.6	240.13
1987	6.6	242.29
1988	6.2	244.5
1989	6.5	246.82
1990	6.2	249.62
1991	7.2	252.98
1992	7.4	256.51
1993	7.8	259.92
1994	7.7	263.13
1995	8.7	266.28
1996	7.6	269.39
1997	10.1	272.65
1998	10.5	275.85
1999	10.9	279.04
2000	12.1	282.16
2001	13.1	284.97
2002	13.5	287.63
2003	14.1	290.11
2004	15.2	292.81
2005	16.3	295.52
2006	17.3	298.38
2007	17.4	301.23
2008	18.8	304.09
2009	20.7	306.77
2010	21.1	308.11
2011	20.7	310.45
2012	21.5	312.76
2013	22.3	314.96
2014	22	317.34

Table 2: The U.S. Diabetic population graph and the graph information including the graph type, equation of the graph and the error value.

Diabetes Population Growth

Cubic

$y = .000308x^3 + 1.86x^2 - 3755.99x + 2.52E{+}06$

Error Value: 3.709

Diabetes Population Growth

Quadric

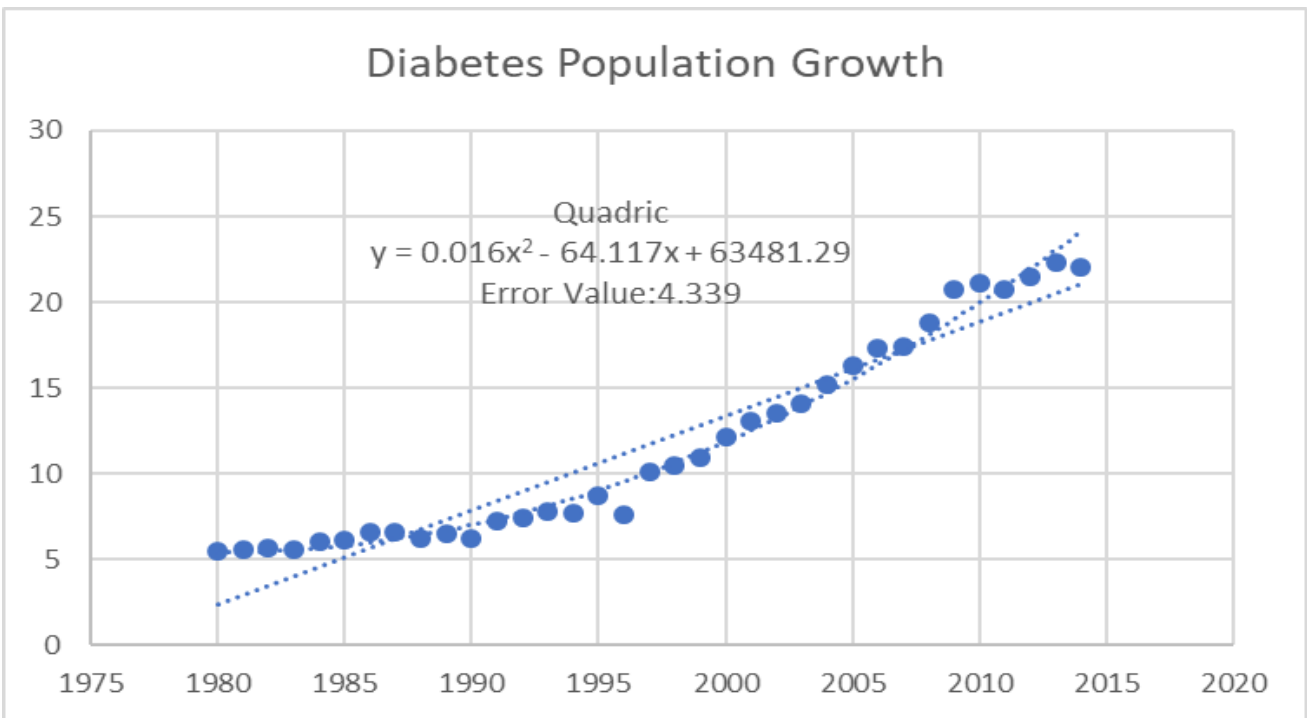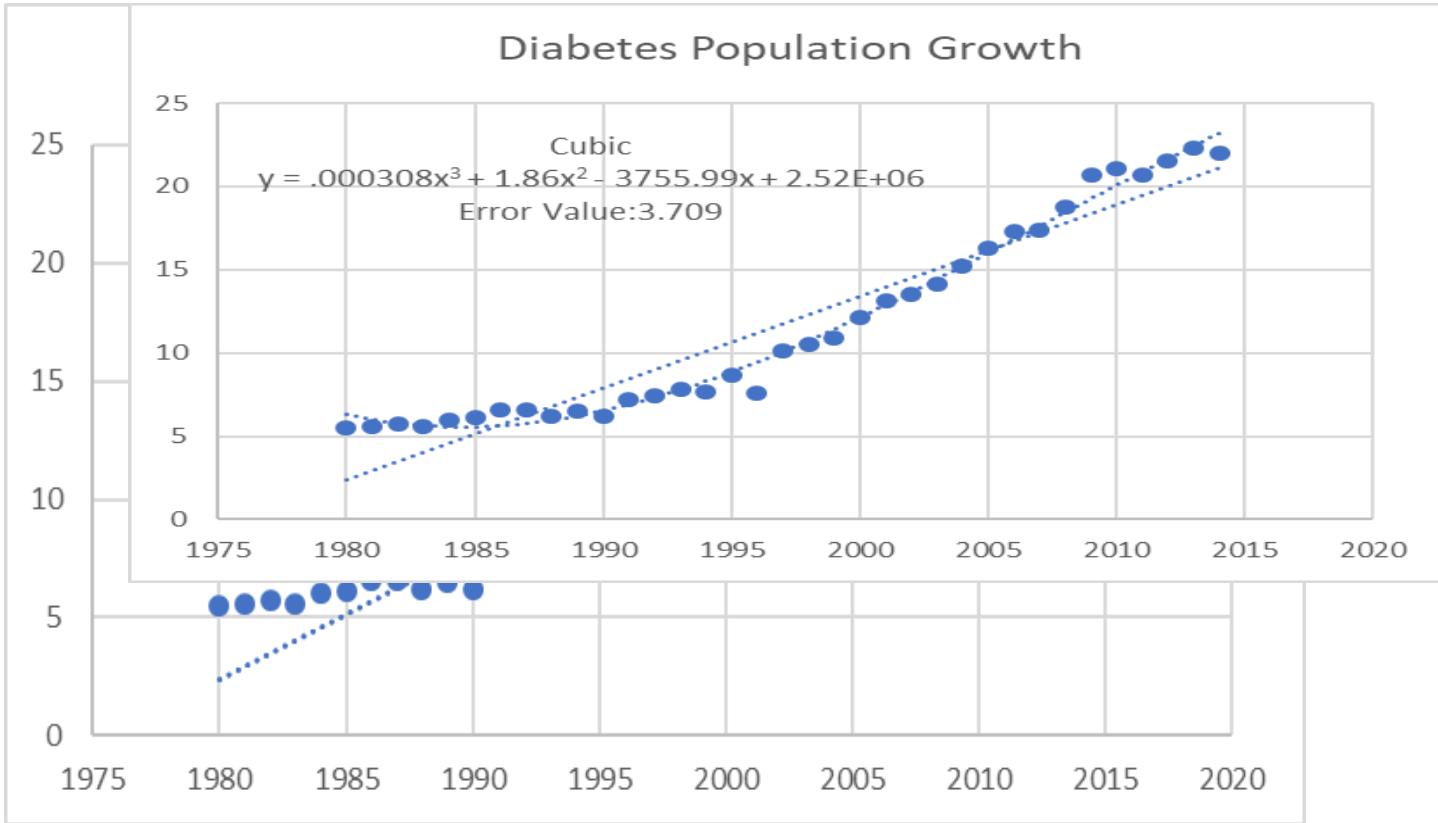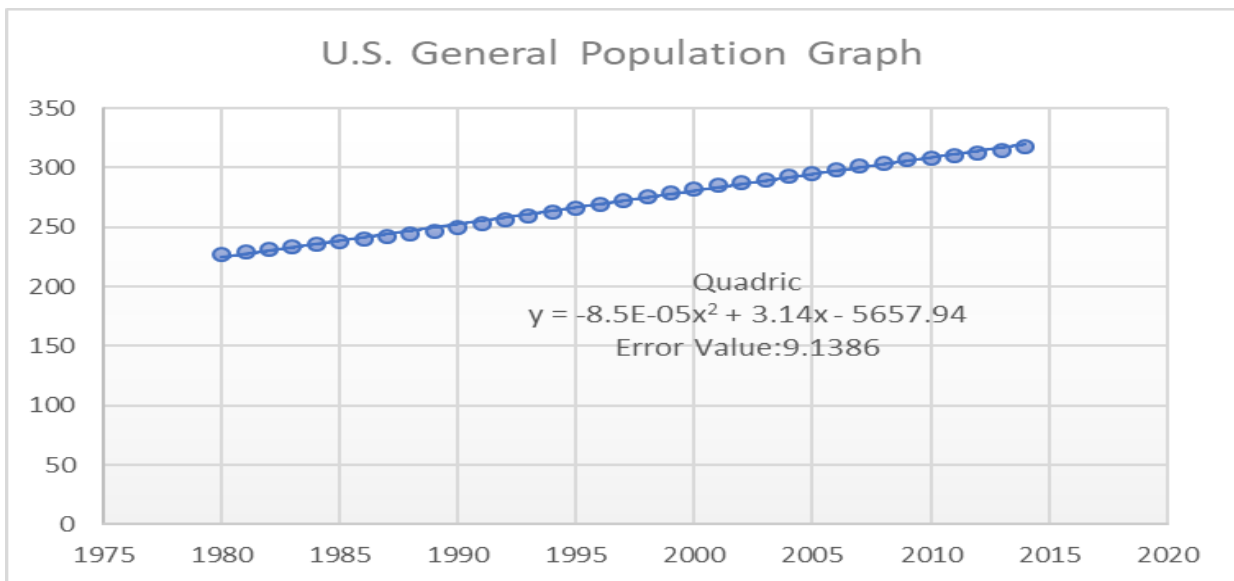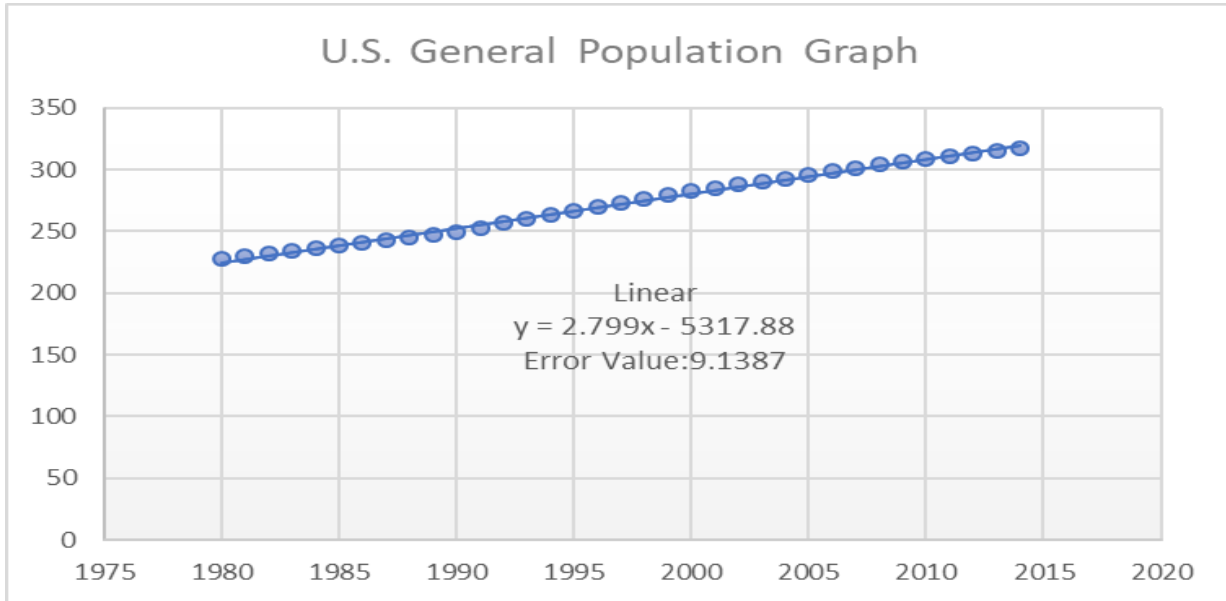$y = 0.016x^2 - 64.117x + 63481.29$

Error Value: 4.339

Table 3: The U.S. general population graph and the graph information including the graph type, equation of the graph, and error value.

U.S. General Population Graph

Cubic
$$y = .000625x^3 + 3.75x^2 - 7478.66x + 4.97E+06$$
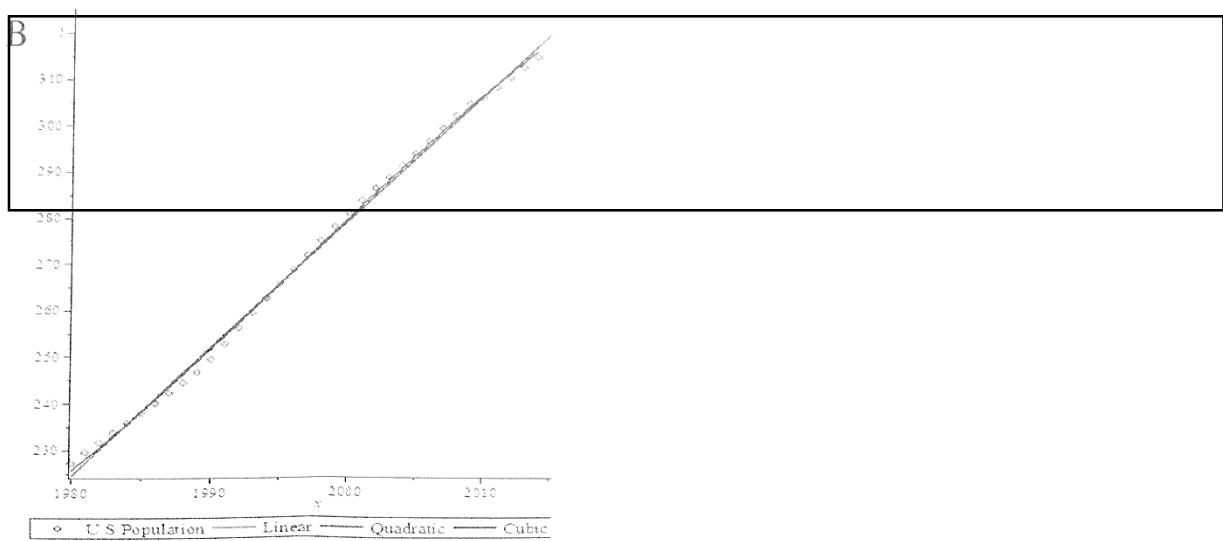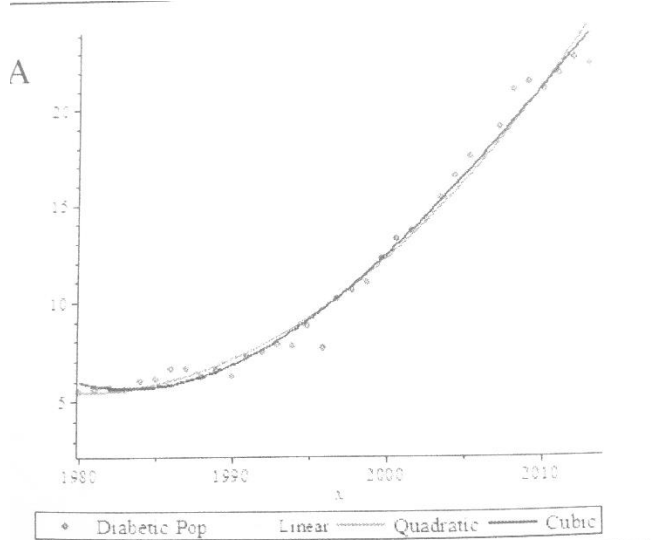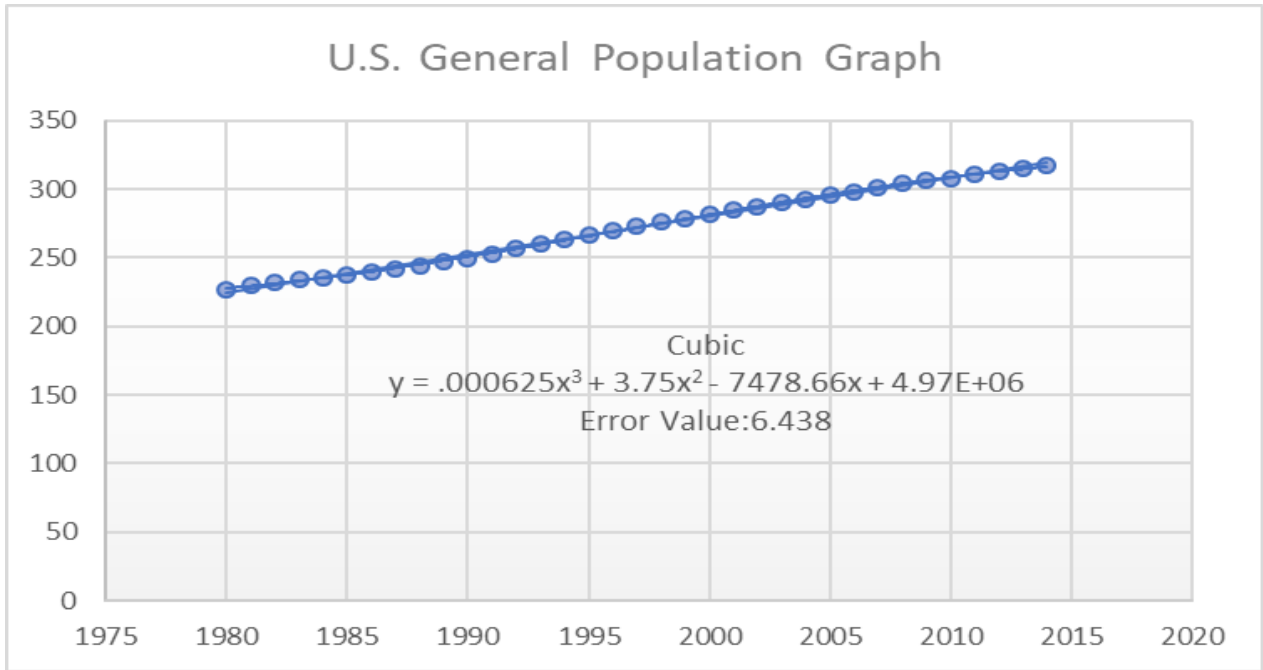Error Value: 6.438

A

B

Figure 3: Comparison of results including prediction along the best.

A: U.S. Diabetic Population

B: U.S. General Population

References

1. Ryan, L (2015), America has the Highest rate of diabetes in the developing world. (*Daily Mail*)
2. Burden, R.L. (2005), *Numerical Analysis* (*9th Edition*). Thomas Brooks: Cole.
3. http://www.census.gov
4. http://www.cdc.gov/diabetes/statistics/prev/national/figraceethsex.htm