IJSEAS

International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-7,July 2016
ISSN: 2395-3470
www.ijseas.com

Statistical Results in Epidemiology Case Study: AIDS

Sara Kasiri Bidhendi
IBO IT, Math and Computer Teacher
Tehran International School, Tehran, Iran

Akram Madadi
Tehran Education Office (R5), Organization of Education, Ministry of Education, , Iran.

Abstract

This paper has allowed us to explore the links between these two fields, broadening our view of mathematics. By observing and analyzing trends in such data, it is possible to perceive which precautionary measures, treatments and medications etc. have been particularly effective in controlling a disease and causing medical advancement. Moreover, we have learned the importance of diagrammatic representation for making the collected data more comprehensible. Different types of graphs can serve as a great visual aid for spotting trends. By studying different factors attributed to the disease (such as its distribution among different locations, age groups, racial groups, genders etc.), it is possible to plan more carefully and appropriately for the future. Health care plans can be adjusted to become more effective, targeting the more vulnerable groups. Furthermore, the disease can be monitored much more effectively and sudden outbreaks of the disease can be prevented. Although Our exploration focused on investigating and preventing AIDS in one country, we've realized how statistics can be used not only to improve a nation's health situation, but also the global health condition. In this way, mathematics even saves lives!

Introduction:

Ever since we were introduced to the mathematical exploration, WE was sure that we wanted to explore a topic involved in the medical field. As we plan on studying medicine in the future, we saw this as the perfect opportunity for me to do research in my field of interest. Seeing as statistics is one of my favorite sections of the syllabus, it became clear that studying medical statistics would be the best choice for me; we chose epidemiology as my main topic. Epidemiology is "a branch of medical science that deals with the incidence, distribution, and control of disease in a population".[1] However, this discipline quite vast, and so it was necessary to further narrow down the scope of the investigation. As we were researching, we came across the issue of AIDS.

AIDS, or Acquired Immune Deficiency Syndrome, is a chronic medical condition that is caused by HIV (the Human Immunodeficiency Virus). The virus attacks the T-cells or CD4-cells (which regulate immune responses) in the immune system. The virus invades these cells to make more copies of itself, destroying the T-cell in the process. Over time, the virus greatly reduces the number of these cells present in the body, resulting in the degradation of the immune system. The immune system gradually weakens and the body's ability to fight against diseases is greatly reduced. When the number of T-cells in the body drops below a specific level, the person is said to have AIDS. With a terribly damaged

[1] "Epidemiology." Merriam-Webster. Accessed February 13, 2015. http://www.merriam-webster.com/dictionary/epidemiology.

immune system, people with AIDS are vulnerable to opportunistic infections such as pneumonia, various parasitic/viral/fungal infections and even some types of cancer, which may result in their death. A person infected with HIV is infected for life. Currently, there is no cure for AIDS; there are only treatments which can suppress the HIV virus or somewhat strengthen the immune system.[2]

We discovered that since the beginning of the AIDS epidemic in the early 1980s, over 78 million people have been infected with HIV, and this number continues to grow today. we also found out that AIDS is one of the leading causes of death.[3] we was intrigued to pursue this topic as it not only involved an interesting subject matter, but would also allow me to learn more about medicine and AIDS from the point of view of a math student. we first wanted to track the spread of AIDS over the years, and so we required a large database; we decided to focus on AIDS in the US, as there would be more data at my disposal. we found a substantial archive of HIV/AIDS surveillance reports for the US ranging from 1982-2012; however, combining the reports would lead to inaccurate results as some editions had revised the statistics already existing in previous editions. Hence, we decided to use only two editions: the year-end 2001 edition and the 2012 edition. In the first part of my exploration, we will use the data from the 2001 report to study the spread of AIDS in the past from the emergence of the disease until the beginning of the 21st century in order to observe any trends and better understand the spread of this disease. In the second section of this investigation, we will use the 2012 report to look at more recent data and focus on controlling and preventing AIDS in the future.

It is worth noting that out of the American population, the adults and adolescents (≥13 years of age) were taken as the sample to be investigated in this study. Furthermore, “people” and “Americans” have been used to refer to the adult population living in the US.

The Number of Americans Living with AIDS:

In epidemiology, *prevalence* is the number of people who have the disease at a given point in time.[4] we began the investigation by first examining the number of Americans living with AIDS over a time interval, or the prevalence of AIDS in the US. In this way, we could get an overall idea of the burden of AIDS among the American population. Time was chosen as the independent variable to be investigated and the number of Americans living with AIDS was taken to be the dependent variable. However, due to insufficient available data, the maximum time interval that could be chosen was from the year 1993 to 2001. Table 1 shows the data that was considered.[5] It is worth noting that these values do not actually represent the precise number of people living with AIDS during these years; they are estimates which have been calculated (in the report

[2] "What Is HIV/AIDS?" AIDS.gov. April 29, 2014. Accessed February 13, 2015. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/what-is-hiv-aids/.

[3] "Global HIV & AIDS Epidemic." AVERT. December 12, 2014. Accessed February 13, 2015. http://www.avert.org/global-hiv-aids-epidemic.htm.

[4] "Descriptive Epidemiology: Rates, Ratios, Etc." Ottawa. November 11, 2014. Accessed February 13, 2015. http://www.med.uottawa.ca/SIM/data/Epidemiology_rates_e.htm.

[5] "U.S. HIV and AIDS Cases Reported through December 2001." *HIV/AIDS Surveillance Report* 13, no. 2 (2001): 36. Accessed February 6, 2015. http://www.cdc.gov/hiv/pdf/statistics_2001_HIV_Surveillance_Report_vol_13_no2.pdf.

itself) by taking the difference between the estimated cumulative number of people diagnosed with AIDS and the estimated cumulative number of deaths of people with AIDS. [6] Throughout this exploration, statistically adjusted (estimated) values which accounted for reporting delays were used in order to arrive at more accurate results. The values in Table 1 were used to create a scatter plot was created in order to see if there were any trends. In addition, a line of best fit or regression line was drawn to allow for interpretation and prediction purposes. Graph 1 shows the scatter plot, which was made using Microsoft Office Excel 2007.

Pearson product-moment correlation coefficient, *r*.[7]

$$r = \frac{S_{xy}}{S_x S_y}$$

where

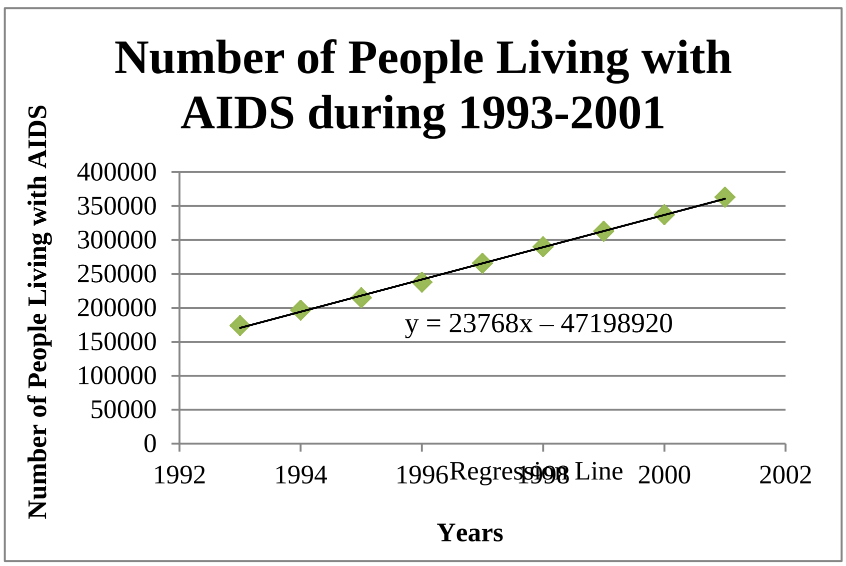
$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$S_x = \sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$S_y = \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}$$

Table 1: The number of people living with AIDS in America in 1993-2001.

Year	Number of People Living with AIDS
1993	173,772
1994	196,452
1995	214,711
1996	237,735
1997	265,464
1998	289,709
1999	312,804
2000	337,017
2001	362,827



Graph 1: A scatter plot showing the number of people living with AIDS in America from 1993 to 2001.

The data suggests that a strong, positive, linear correlation exists between the two variables: as time increases, the number of Americans living with AIDS also increases. Since there is a linear relationship between *x* and *y*, the correlation of these two variables can be calculated numerically using the

In the above equations, *x* is the time (in years), *y* is the number of people living in the US with AIDS and *n* is the number of values (9). Using LinRegTTest, the GDC returned a value of +0.999 (rounded to 3

[6] Ibid.

[7] Buchanan, Laurie, Jim Fensom, Ed Kemp, Paul La Rondie, and Jill Stevens. "Bivariate Analysis." In *Mathematics Standard Level*, 350. Oxford: Oxford University Press, 2012.

decimal places) for the *r*-value. This is shown in Figure 1 below. Thus, the correlation between the years and the number of Americans living with AIDS is strong, suggesting that it is possible to use the graph to make predictions; however, first an equation must be determined for the line of best fit.

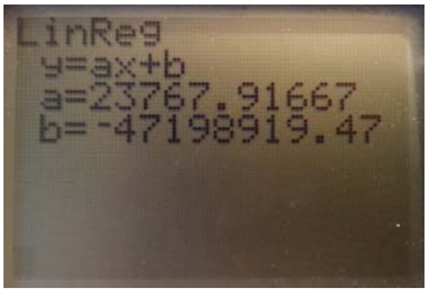


Figure 2: A GDC was used to fine the equation of the line of regression.

Since the correlation is linear, the equation for the line of regression is in the form of $y = ax + b$. Using a TI-84 Texas Instruments GDC *(see Figure 2)*, the equation of the regression line was found to be:

$$y = 23768x - 47198920$$

where x is the time in years and y is the number of people in the US living with AIDS. Note that the values of a and b have been rounded up to the nearest integer since the number of people are being considered. The slope of this line is equal to a, 23768, and it can be interpreted as the number of people living with AIDS per year. The data suggests that every year, there are 23,768 more people living with AIDS in the United States than the previous year. However, the equation obtained is only valid when $x \geq 1985$. When a year that is less than 1985 is substituted into the equation, a negative number is returned for the y value:

$$y = 23768(1984) - 47198920$$

$$y = -43208$$

This is incorrect as the number of people cannot be negative. Furthermore, the y-intercept of this scatter plot doesn't allow for an appropriate interpretation. It proposes that the number of Americans living with AIDS (when $x = 0$) is – 47198920, and a negative number in this case is unsuitable.

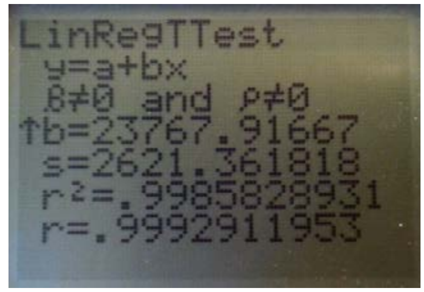


Figure 1: A GDC was used to determine the *r-value*.

As mentioned earlier the high *r*-value (0.999) attained suggests that the data can be used to make predictions. For example, in the year 2011, the number of people living with AIDS in the US is:

$$y = 23768(2011) - 47198920$$

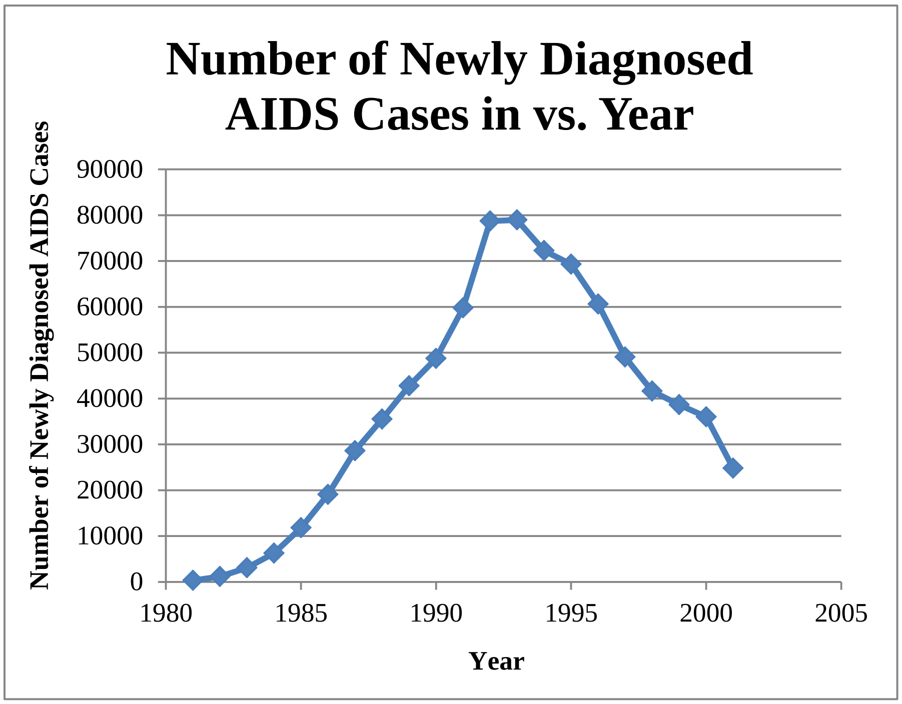
$$y = 598528$$

That is, in the year 2011, there are 598,528 people living with AIDS in the US. The year 2011 was chosen purposefully in order to compare it to the actual value that exists, as reported in the 2012 HIV/AIDS surveillance report (the latest published report). According to this report, the estimated number of people living with AIDS in the US in 2011 is 494,602.[8] Hence, line of regression is not accurate for making predictions as it overestimated the number of people.

[8] "Diagnoses of HIV Infection in the United States and Dependent Areas, 2012." *HIV Surveillance Report* 24 (2014): 55-6. Accessed February 4, 2015. http://www.cdc.gov/hiv/pdf/statistics_2012_HIV_Surveillance_Report_vol_24.pdf.

As aforementioned, these values represent the prevalence of AIDS in the US in 1993-2001. Although they present one with an overall idea of the magnitude of the population infected with AIDS, they give no indication of when the disease was contracted. Since the number of people living with AIDS is calculated by subtracting the number of deaths from the number of newly diagnosed, we decided to investigate these two contributing factors separately over a specific time interval. Therefore, we will look for trends in the spread of AIDS over a twenty year interval (from 1981 to 2001), since this was the greatest time period for which the required data could be gathered. First, a graph will be plotted for the number of people newly diagnosed with AIDS during this time period. Table 1 displays the data which was used to plot the time series shown in Graph 2.[9] Microsoft Office Excel 2007 was used to plot this graph.

At this point, it is worth defining a new term in epidemiology known as *incidence*. Incidence is the rate or number of newly diagnosed cases of a disease in a fixed time period, usually one year.[10] However, the data that will be used has been termed as the "cases diagnosed during interval". According to the 2001 HIV/AIDS surveillance report, the "diagnoses do not necessarily represent new infections (incidence)" because "some persons were



Graph 2: The number of people newly diagnosed with AIDS in America from 1981 to 2001.

[9] "U.S. HIV and AIDS Cases Reported through December 2001." *HIV/AIDS Surveillance Report* 13, no. 2 (2001): 30. Accessed February 6, 2015. http://www.cdc.gov/hiv/pdf/statistics_2001_HIV_Surveillance_Report_vol_13_no2.pdf.

[10] "Descriptive Epidemiology: Rates, Ratios, Etc." UOttawa. November 11, 2014. Accessed February 13, 2015. http://www.med.uottawa.ca/SIM/data/Epidemiology_rates_e.htm.

infected recently, and others were infected at some time in the past".[11] Thus, for the purpose of accuracy, the data shown below has been labeled as the cases of diagnoses rather than 'incidence'. Nevertheless, it does give an approximate impression of the incidence during the years.

Table 2: The number of people newly diagnosed with AIDS in America in 1981-2001.

Year	Number of People Newly Diagnosed with AIDS
1981	323
1982	1,170
1983	3,076
1984	6,247
1985	11,794
1986	19,064
1987	28,599
1988	35,508
1989	42,768
1990	48,732
1991	59,760
1992	78,705
1993	78,954
1994	72,266
1995	69,307
1996	60,613
1997	49,062
1998	41,605
1999	38,640
2000	35,986
2001	24,804

Graph 2 conveys a great deal of information regarding the spread of AIDS. The graph increases on the interval $1981 \leq year \leq 1993$. In fact, AIDS was first discovered in 1981 and it broke out in an epidemic, quickly spreading amongst the population; the graph has a rather steep positive slope at first. The graph reaches its peak in 1993; the greatest number of people was diagnosed with AIDS in this year. The number finally began to decrease in 1994. This is most likely due to the fact that more precautionary measures were being taken and the US population had become more careful with respect to the spread of this disease. In fact, in 1993, the female condom was approved and in the following years, various reports were published regarding the prevention of HIV. As a result, the number of people diagnosed with HIV and hence AIDS decreased.[12] The graph shows that this number continues to decrease on the interval $1994 \leq year \leq 2001$, but with varying slopes.

The second factor affecting the prevalence of AIDS in the US is mortality. Using Excel, another graph was plotted (Graph 3) for the number of deaths in people with AIDS in the US over the same time period considered above (1981-2001). Table 3 displays the data that was used.[13] It should be noted that the deaths are not necessarily caused by HIV/AIDS-related diseases.[14]

[11] "Diagnoses of HIV Infection in the United States and Dependent Areas, 2012." *HIV Surveillance Report* 24 (2014): 6. Accessed February 4, 2015. http://www.cdc.gov/hiv/pdf/statistics_2012_HIV_Surveillance_Report_vol_24.pdf.

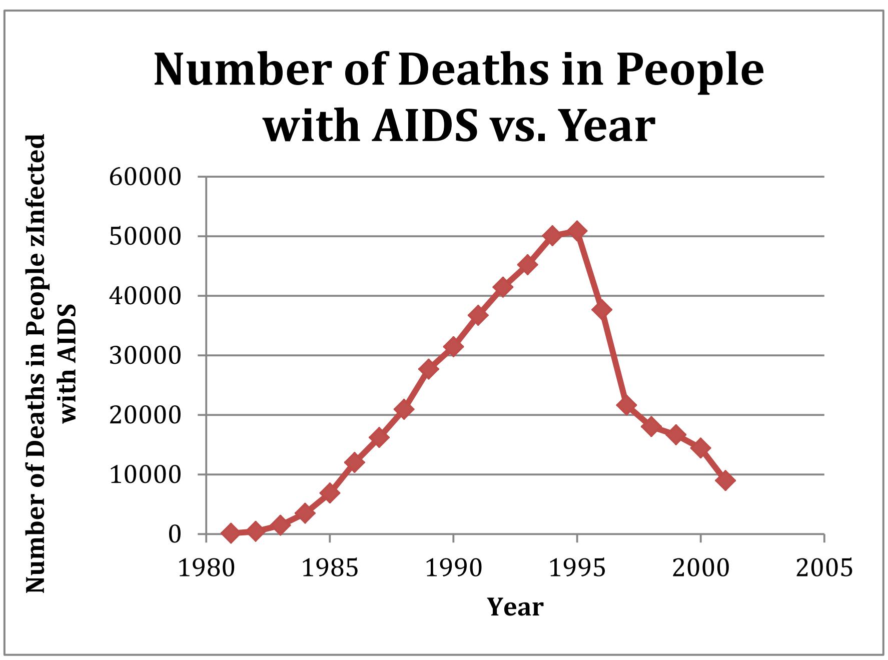
[12] "A Timeline of AIDS." AIDS.gov. Accessed February 13, 2015. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/aids-timeline/index.html.

[13] "U.S. HIV and AIDS Cases Reported through December 2001." *HIV/AIDS Surveillance Report* 13, no. 2 (2001): 30. Accessed February 6, 2015. http://www.cdc.gov/hiv/pdf/statistics_2001_HIV_Surveillance_Report_vol_13_no2.pdf.

[14] Ibid.

Table 3: The estimated cumulative number of people diagnosed with AIDS, the estimated cumulative number of deaths in people with AIDS and the estimated number of people living with AIDS in America in 1981-2001.

Year	Estimated Cumulative Number of People Diagnosed with AIDS	Estimated Cumulative Number of Deaths in People with AIDS	Estimated Number of People Living with AIDS (Prevalence)
1981	323	122	201
1982	1,493	575	918
1983	4,569	2,056	2,513
1984	10,816	5,530	5,286
1985	22,610	12,407	10,203
1986	41,674	24,423	17,251
1987	70,273	40,617	29,656
1988	105,781	61,539	44,242
1989	148,549	89,219	59,330
1990	197,281	120,655	76,626
1991	257,041	157,363	99,678
1992	335,746	198,787	136,959
1993	414,700	243,974	170,726
1994	486,966	294,045	192,921
1995	556,273	344,921	211,352
1996	616,886	382,567	234,319
1997	665,948	404,197	261,751
1998	707,553	422,225	285,328
1999	746,193	438,873	307,320
2000	782,179	453,306	328,873
2001	806,983	462,269	344,714



Graph 3: The number of deaths in people with AIDS in America from 1981 to 2001.

Table 4: The number of deaths in people with AIDS in America in 1981-2001.

Year	Number of Deaths in People with AIDS
1981	122
1982	453
1983	1,481
1984	3,474
1985	6,877
1986	12,016
1987	16,194
1988	20,922
1989	27,680
1990	31,436
1991	36,708
1992	41,424
1993	45,187
1994	50,071
1995	50,876
1996	37,646
1997	21,630
1998	18,028
1999	16,648
2000	14,433
2001	8,963

Graph 3 shows that the number of deaths due to AIDS-related illnesses steadily increased ever since the epidemic began, reaching a peak in 1995. Following this year, there was a sudden drop in the number of deaths. In fact, in 1995, the first protease inhibitor (antiviral drugs used to treat HIV/AIDS) was approved; this marked the beginning of highly active antiretroviral therapy (HAART). [15] HAART involves using a combination of antiviral drugs to aggressively suppress the progression of the HIV infection. In the next following years, the number of deaths due to AIDS continued to decrease as more precautionary measures were taken and more drugs were created and administered (such as NNRTI and nevirapine). [16] The number of deaths continued to fall afterwards but with a less steep slope.

The shape of Graph 3 is quite similar to that of Graph 2; however, these two graphs cannot be compared as is because of their different scales. For effective comparison, we decided to put these two graphs of the same scale. In addition, we will also include a graph for the number of people living with AIDS (i.e. the prevalence) at each year, using the data already presented in Tables 2 and 3. By doing so, we can compare my newly calculated prevalence to that which was presented in Table 1 and Graph 1. Following the format of the report used, we employed the following equation for calculating the prevalence of AIDS at each year:

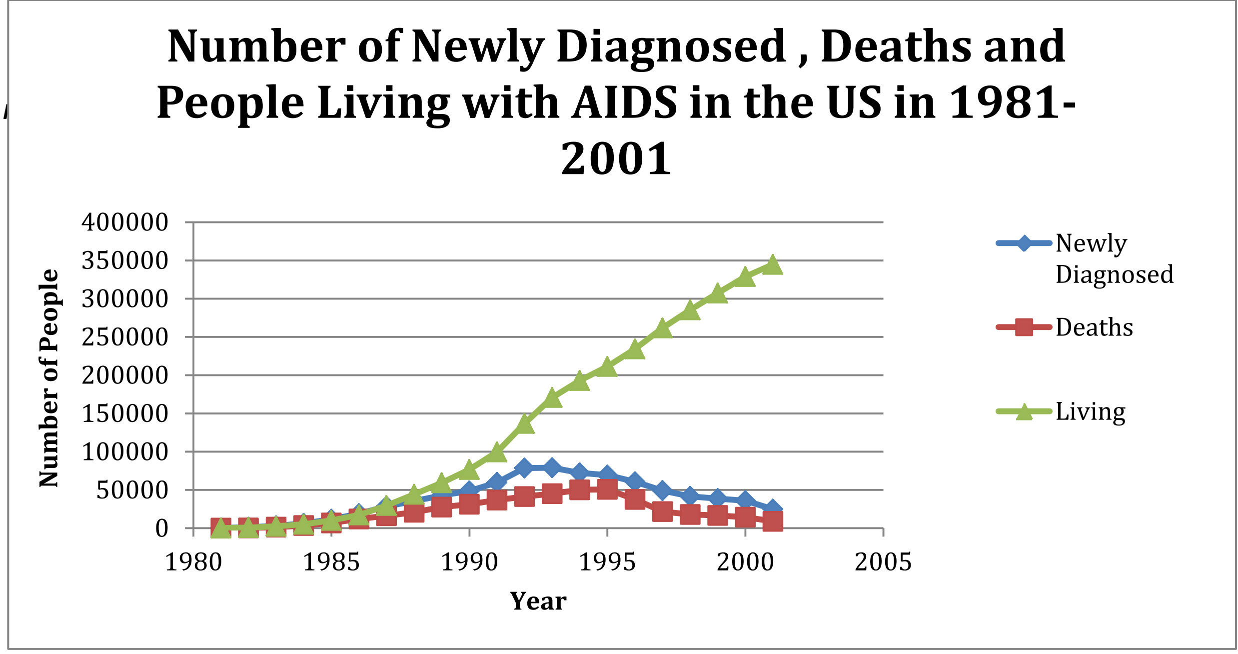
Estimated Number of People Living with AIDS = Estimated Cumulative Number of People Diagnosed with AIDS − Estimated Cumulative Number of Deaths in People with AIDS

Graph 4 shows a combination of Graphs 2 and 3 as well as the newly calculated prevalence for the years 1981-2001.

The calculations were done using Microsoft Office Excel 2007 and are shown in Table 4. First, the cumulative frequency was calculated for the deaths as well as the

[15] "Diseases of the Immune System." In *Robbins and Cotran Pathologic Basis of Disease*, edited by Vinay Kumar, Abul K. Abbas, and Jon C. Aster, 255. 9th ed. Vol. 1. Philadelphia: Elsevier Saunders, 2015.

[16] "A Timeline of AIDS." AIDS.gov. Accessed February 13, 2015. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/aids-timeline/index.html.



Graph 4: The number of newly diagnosed, deaths, and people living with AIDS in the US in 1981-2001.

number of people who were diagnosed in each year. The difference between these two values was then taken in order to find the estimated number of people living with AIDS in each year.

Several features of this graph are worth noting. As aforementioned, the number of people diagnosed began decreasing only after 1993, whereas the number of deaths began decreasing after 1995. Between these two years, the number of people diagnosed with AIDS was decreasing; however, the number of AIDS-related deaths continued to increase simultaneously. This is because the public health of the nation was improving: new measures were being taken which reduced the number of people becoming infected with HIV and hence AIDS. However, the deaths continued to steadily rise in this period because no effective measures had been taken to treat AIDS itself. The deaths decreased only after the introduction of HAART in 1995.

Furthermore, it can be seen that the prevalence of AIDS steadily increased over this time period. However, this increase is not perfectly linear. Graph 1 suggested that the correlation between the number of people living with AIDS and time was linear. In the graph above, the curve is exponential at first (until approximately the year 1992), and then it becomes a noticeably straight line. Therefore, it was wrong to extrapolate graph 1, as the prevalence was linear for 1993-2001, not the other years.

However, it is worth mentioning that the newly calculated prevalence values shown in Table 4 and Graph 4 are all lower than those given by the report, as shown in Table 1. This is probably because the report itself adjusted the values due to reporting delays and other issues encountered in collecting the data.

Now that the spread of AIDS has been studied for the past years, the rest of this exploration will focus on how to prevent AIDS in the future. As AIDS is originally caused by HIV, we will focus on the prevention of the spread of HIV (which ultimately prevents the spread of AIDS). Thus, the diagnoses of HIV will be studied in this section rather than the diagnoses of AIDS.

Also, from here on, the 2001 HIV/AIDS surveillance report was no longer used, but rather focus will be turned to the latest edition, the 2012 report.

Table 5: The frequency and cumulative frequency for the number of people diagnosed with AIDS between the ages of 13 and 65.

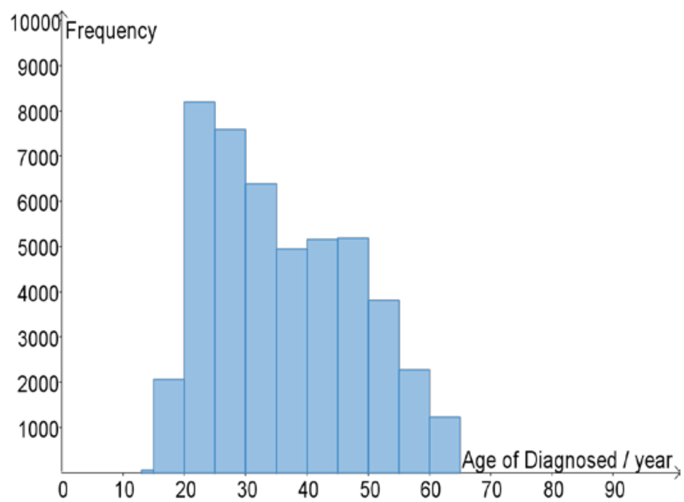
The Number of Americans Diagnosed with HIV in 2012

To begin with, we will first analyses the diagnoses of HIV among different age groups. Age has been considered as a continuous data type, and so a histogram has been drawn. Since a histogram will be drawn, the lower and upper boundaries of the data must be defined for each class interval. Hence the ages <13 have not been considered, as there's no defined lower limit; ages ≥ 65 have also not been considered, as there is no defined upper limit. As a result, the data includes an age **range** of 52: a minimum age of 13 and a maximum age of 64. Let a be the age of the infected people. Table 1 is a grouped frequency table that shows the data that was used as well as the midpoint age and cumulative frequency that were calculated using Excel.[17] Note that frequency is the number of people diagnosed with HIV. The midpoint age was found using the following formula:

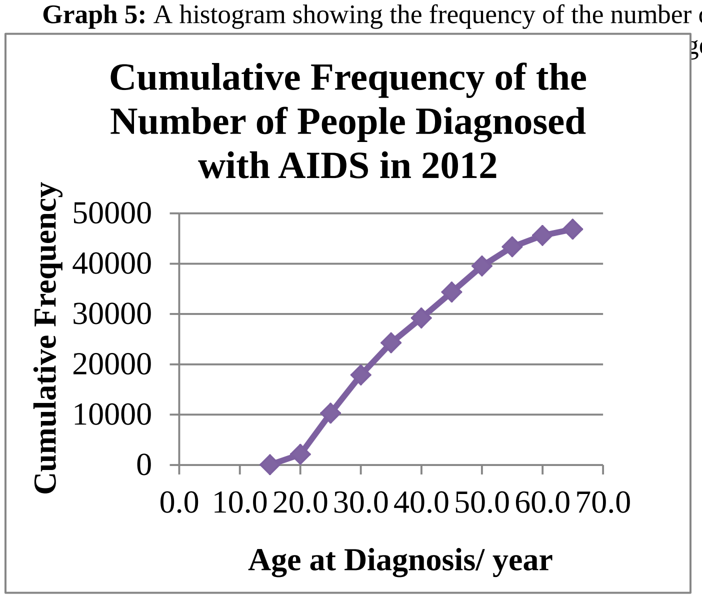
$$\text{Midpoint Age} = \frac{\text{lower bound} + \text{upper bound}}{2}$$

The data in Table 5 was used to create the histogram shown in Graph 5 and the cumulative frequency graph shown in Graph 6. Geogebra was used to make Graph 5 and Excel was used to make Graph 6.

Age at Diagnosis (year)	**Midpoint Age (year)**	**Frequency**	**Cumulative Frequency**
$13 \leq a < 15$	14.0	51	51
$15 \leq a < 20$	17.5	2,053	2,104
$20 \leq a < 25$	22.5	8,187	10,291
$25 \leq a < 30$	27.5	7,589	17,880
$30 \leq a < 35$	32.5	6,388	24,268
$35 \leq a < 40$	37.5	4,939	29,207
$40 \leq a < 45$	42.5	5,145	34,352
$45 \leq a < 50$	47.5	5,183	39,535
$50 \leq a < 55$	52.5	3,800	43,335
$55 \leq a < 60$	57.5	2,269	45,604
$60 \leq a < 65$	62.5	1,221	46,825



Graph 5: A histogram showing the frequency of the number of ...ges.



Graph 6: A graph showing the cumulative frequency of the

[17] "Diagnoses of HIV Infection in the United States and Dependent Areas, 2012." *HIV Surveillance Report* 24 (2014): 26. Accessed February 4, 2015. http://www.cdc.gov/hiv/pdf/statistics_2012_HIV_Surveillance_Report_vol_24.pdf.

International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-7,July 2016
ISSN: 2395-3470
www.ijseas.com

It can be seen from Graphs 5 and 6 that the majority of Americans diagnosed with HIV in 2012 were between 20 and 55 years old. In the histogram, the highest frequencies belong to these ages, and, the graph steadily increases but is more flat for the youngest and oldest population.

Furthermore, the above data has the following measures of central tendency:

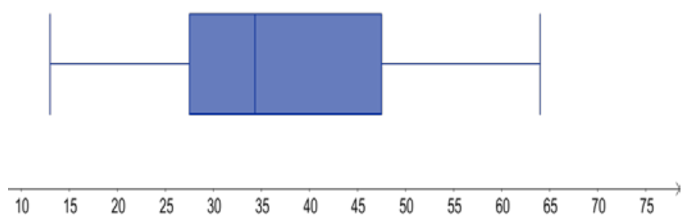
- The **modal class** is $20 \leq a < 25$. → This means that age interval 20-24 years includes the greatest number of people infected with HIV (i.e. it has the highest frequency).
- The **median** is equal to 34.33. → 50% of the infected people have an age greater than 34.33 years and 50% have an age lower than 34.33 years.
- The **mean** is 36.17. → The average American infected with HIV has an age of 36.17 years.
-

These were calculated using Geogebra. Note that since intervals are being considered for the *x* values, the mean can be determined using the following formula: Mean $= \frac{\sum fm}{\sum f}$

where *f* is the frequency or number of people with AIDS and *m* is the midpoint for each age interval (shown in Table 5). Note that it was assumed that all of the infected people were equally spread around the mid age. In addition, the above data has the following measures of dispersion:

- The **first quartile**, Q_1, is 27.5. → 25% of the sick people are 13-27.5 years old.
- The **third quartile**, Q_3, is 47.5. → 75% of the sick people are 47.5-64 years old.
- The **interquartile range**, *IQR*, is 20. → The middle 50% of the sick people are between 27.5 and 47.5 years old.

Note that the measures of central tendency and the measures of dispersion shown above were calculated using Geogebra *(see Figure 3)*.The values calculated above were then used to create a box and whisker plot, shown in Graph 7 below.

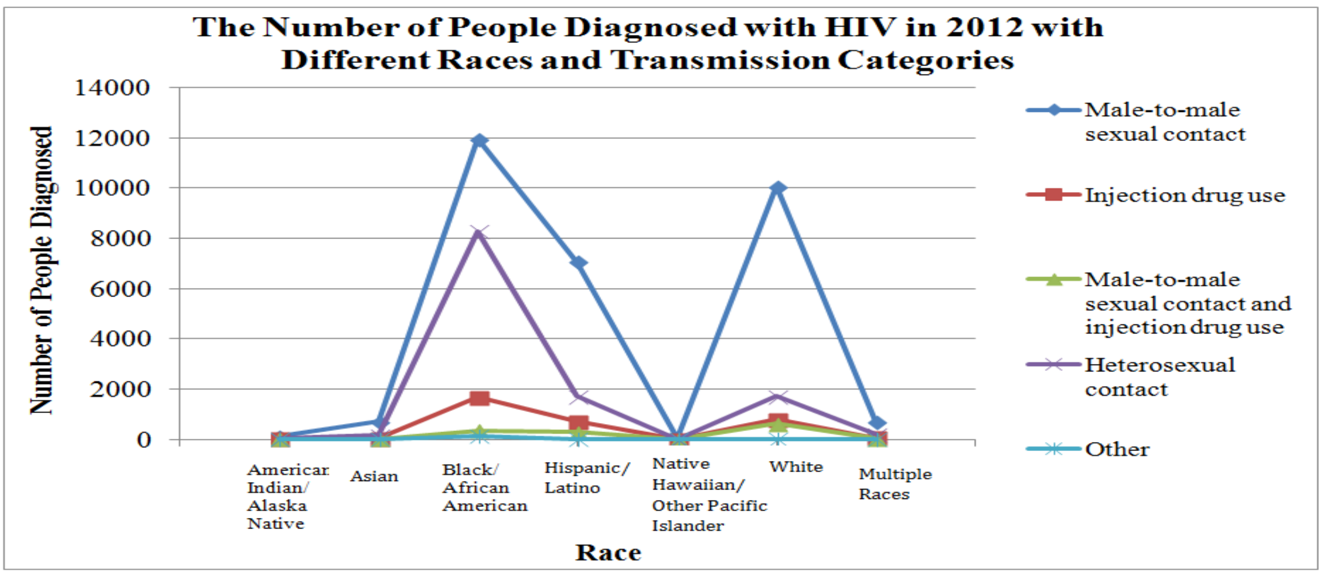


Graph 7: A box and whisker plot of the people in America diagnosed with HIV at different ages in

- The **variance**, σ^2, is 143.07.
- The **standard deviation**, σ, is 11.96 years.

Table 6: The number of people diagnosed with HIV in 2012 with respect to their race and transmission category.

Transmission Category	Races						
	American Indian/ Alaska Native	Asian	Black/ African American	Hispanic/ Latino	Native Hawaiian/ Other Pacific Islander	White	Multiple Races
Male-to-male sexual contact	132	720	11,957	7,057	58	10,069	711
Injection drug use	28	32	1,682	704	3	793	69
Male-to-male sexual contact and injection drug use	19	16	352	285	7	636	43
Heterosexual contact	48	183	8,282	1,733	11	1,733	201
Other	0	2	131	26	0	28	2



Graph 8: A line graph showing the number of people infected with HIV in 2012 with respect to their race and transmission category.

To further investigate how the diagnosis of HIV was spread among the infected, two other factors were considered: their races and transmission category (i.e. how they acquired HIV). Table 6 shows the data that was used.[18] This data was used to create the line graph shown in Graph 8, which was constructed using Excel.

[18] Ibid.

Graph 8 allows for a multivariate analysis of the people diagnosed with HIV. Generally, black/African Americans and white races experienced the greatest diagnosis of HIV, as shown by the two "peaks" in this graph. Hispanics/Latinos also suffered from a relatively larger number of HIV diagnoses. The other races did not have as many infected, and the data points are quite low on the graph. In addition, it can be seen that the most common form of HIV transmission is male-to-male sexual contact, regardless of race (the blue line has the highest y-values for all x-values). The next most common transmission route by which HIV was contracted was heterosexual contact (the purple line). Again, for both of these transmission categories, the people diagnosed were mainly black/African American or white.

Therefore, in the future, measures to prevent the spread of HIV should involve controlling particularly male-to-male sexual contact and heterosexual contact. In addition, the US should focus on preventing the spread of HIV specifically in Hispanic/Latino groups and more notably, black/African Americans and white racial groups. These are vulnerable groups that are more susceptible to becoming diagnosed with HIV. For instance, prevention of HIV may be in the form of informative advertisements warning the population of HIV/AIDS.

Furthermore, from Graphs 5, it can be seen that the age interval 20-24 years includes the greatest number of people diagnosed with HIV (the modal class). That is, this is the age group most vulnerable to becoming diagnosed; thus, the US should focus on preventing the spread of HIV particularly for these ages.

Conclusion

Even though this exploration was successful for the most part, it did encounter several drawbacks. The values used were only estimates; although they were successful in conveying the general trends, they were not exact. In addition, although this investigation focused on HIV/AIDS in the US, AIDS is a global issue, which affects the lives of many around the world.

Over the course of this exploration, we have come to realize the importance of statistics and more generally mathematics in medicine. This papr has allowed me to explore the links between these two fields, broadening my view of mathematics. we now understand the true importance for the collaboration between the sciences and mathematics. Statistics allows us to track analyses the course of diseases and also medical advancements over time. By observing and analyzing trends in such data, it is possible to perceive which precautionary measures, treatments and medications etc. have been particularly effective in controlling a disease and causing medical advancement. Moreover, we have learned the importance of diagrammatic representation for making the collected data more comprehensible. Different types of graphs can serve as a great visual aid for spotting trends.
Statistics not only enables us to understand the spread of diseases in the past, but also greatly aids us in planning for the future. By studying different factors attributed to the disease (such as its distribution among different locations, age groups, racial groups, genders etc.), it is possible to plan more carefully and appropriately for the future. Health care plans can be adjusted to become more effective, targeting the more vulnerable groups. Furthermore, the disease can be monitored much more effectively and

sudden outbreaks of the disease can be prevented.

Although my exploration focused on investigating and preventing AIDS in one country, I've realized how statistics can be used not only to improve a nation's health situation, but also the global health condition. In this way, mathematics even saves lives!

Bibliography

[1] "A Timeline of AIDS." AIDS.gov. Accessed February 13, 2015. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/aids-timeline/index.html.

[2] Buchanan, Laurie, Jim Fensom, Ed Kemp, Paul La Rondie, and Jill Stevens. "Bivariate Analysis." In *Mathematics Standard Level*, 350. Oxford: Oxford University Press, 2012.

[3] "Descriptive Epidemiology: Rates, Ratios, Etc." UOttawa. November 11, 2014. Accessed February 13, 2015. http://www.med.uottawa.ca/SIM/data/Epidemiology_rates_e.htm.

[4] "Diagnoses of HIV Infection in the United States and Dependent Areas, 2012." *HIV Surveillance Report* 24 (2014). Accessed February 4, 2015. http://www.cdc.gov/hiv/pdf/statistics_2012_HIV_Surveillance_Report_vol_24.pdf.

[5] "Diseases of the Immune System." In *Robbins and Cotran Pathologic Basis of Disease*, edited by Vinay Kumar, Abul K. Abbas, and Jon C. Aster, 255. 9th ed. Vol. 1. Philadelphia: Elsevier Saunders, 2015.

[6] "Epidemiology." Merriam-Webster. Accessed February 13, 2015. http://www.merriam-webster.com/dictionary/epidemiology.

[7] "Global HIV & AIDS Epidemic." AVERT. December 12, 2014. Accessed February 13, 2015. http://www.avert.org/global-hiv-aids-epidemic.htm.

[8] "U.S. HIV and AIDS Cases Reported through December 2001." *HIV/AIDS Surveillance Report* 13, no. 2 (2001). Accessed February 6, 2015. http://www.cdc.gov/hiv/pdf/statistics_2001_HIV_Surveillance_Report_vol_13_no2.pdf.

[9] "What Is HIV/AIDS?" AIDS.gov. April 29, 2014. Accessed February 13, 2015. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/what-is-hiv-aids/.