Probability And Statistics

You deserve to breathe clean

Team:" game of statistics"

Name	SEC	BN
Nesma Abd El Kader Shafie	2	29
Sara Gamal Gerges	1	19
Eman Ibrahem El badwehy	1	14
Nancy Ayman	2	28
Yara Hisham Shaaban	2	32
Mennatallah Ahmed	2	26

Submitted to Prof. Maha Amin Hassanein
MTH2253Fall, 2022

Contents

Abstract

Air pollution is one of the major environmental issues. It can cause adverse health effects such as cancer, cardiovascular diseases, and high mortality rates. In extreme cases, it can even cause death. It also has a negative impact on the environment it can damage crops and trees in a variety of ways. The main air pollutants of concern are suspended particles and carbon monoxide. Concerning the sources of air pollution, high population density in cities and urbanized areas is a huge contributory factor to air pollution. Air pollution can also arises from transport, local heating, burning fossil fuels, vehicle exhaust fumes, and emissions from agriculture and industry. Although taking measures to diminish air pollution is necessary, multiple challenges must be addressed first. In this article, we used descriptive data and observations to analyze the problem and try to find out solutions to reduce polluting emissions and improve air quality. Using statistical analysis and calculations, predictions were made as to estimate what will happen to air pollution in the future, and the methods of solution we could offer to advance air quality.

Introduction

A-Definition

Air pollution is the contamination of air due to the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials.[1] There are many different types of air pollutants, such as gases (including carbon monoxide, sulfur dioxide, oxides, and carbon dioxide), particulates (both organic and inorganic), and biological molecules. Air pollution can cause diseases, allergies, and even death to humans.

Figure 1 air pollution

B-Impacts

➢ **Harming Human Health**

According to the World Health Organization, an estimated seven million people die each year from air pollution. Air pollution is the 4th largest threat to human health.

Short-term exposure to air pollution can cause:	Long-term exposure to air pollution can cause:
• Coughing	• Respiratory diseases (asthma, emphysema)
• Wheezing/difficulty breathing	• Cardiovascular damage
• Irritation to eyes, nose, and throat	• Harm to liver, spleen, and blood
• Headache	• Nervous system damage
• Dizziness	• Cancer
• Fatigue	• Birth defects
	• Death

Figure 2Effects of Air pollution

➢ **Harming Animals and Plants**

Wildlife can experience many of the same negative health effects of air pollution that humans do. Damage to respiratory systems is the most common effect on animals, but neurological problems and skin irritations are also common. The air pollutants creating smog have been found in the lungs of dolphins, causing black lung disease (Dyson, 2021).

Plants and crops grow less when exposed to long-term air pollution. Ozone pollution harms plants by damaging structures called stomata- tiny pores on the underside of leaves that allow the plant to breathe-. Between 1980 and 2011, nine billion dollars-worth of soybeans and corn were lost in the US because of ozone pollution.

➢ **Causing Acid Rain**

Burning fossil fuels releases sulfur and nitrogen oxides into the atmosphere. Acid rain forms when sulfur dioxide and nitrogen dioxide mix with water droplets in the atmosphere to make sulfuric acid and nitric acid. Winds can carry these pollutants for thousands of miles, until they fall to the Earth's surface as acid rain, which damages the leaves of vegetation, increases the acidity of soils and water, and is linked to over 500 deaths each year.

➢ **Reducing Sunlight**

A high level of particulate pollution from all types of burning reduces the amount of sunlight that reaches the surface and even changes the appearance of the sky. When less sunlight is available for photosynthesis, forests grow at a slower rate and crops are less productive. Hazy skies not only reduce visibility, but also impact the weather and even the climate.

- **Making a Hole in the Ozone Layer**

The hole in the ozone layer is caused by air pollutants. Chemicals used as refrigerants, such as chlorofluorocarbons (cfcs), contain chlorine atoms. Releasing chlorine atoms into the atmosphere destroys ozone. A single chlorine atom can destroy thousands of ozone molecules.

- **Adding Too Much Nitrogen to the Land**

Gaseous ammonia (NH3) from agriculture and nitrogen dioxide (NO2) from car, truck, and airplane emissions increase the amount of nitrogen in soils. Plants need nitrogen to grow, but too much nitrogen can limit the growth of some plants and increase the growth of others, disrupting the balance of species within an ecosystem. This disruption is negatively impacting grasslands and other fragile environments around the world.

- **Effects of Greenhouse Gas Pollution**

Greenhouse gas pollution is causing climate change. As a result, ecosystems are changing faster than plants and animals can adapt, and many species are going extinct. Marine ecosystems are vulnerable to ocean acidification caused when carbon dioxide emitted into the atmosphere is dissolved in seawater. Ocean acidification makes it difficult for many marine species to grow shells and skeletons.

Melting ice sheets, warming oceans, and extreme weather conditions are examples of how climate changes caused by greenhouse gas pollution threaten ecosystems across the Earth. In many cases, the decline of one or a few species due to air pollution can topple the balance of entire ecosystems.

C-The Importance of fighting air pollution

- **Air pollution control helps to protect human health:** To every human on earth, health is wealth; health is paramount for the running of our day-to-day activities, without which we would have to rely on others to live. Air Pollution Scrubber Suppliers come to help protect our health, which is very valuable.
- **Air pollution control helps prevent economic waste:** With air pollution control, the waste accrued from dead crops and bad water will be limited or stopped. Some California wine producers complained about how their crops were polluted and customers complained that the wine had begun to taste like sulfur. With the presence of air pollution control, economic slowdowns like this will be prevented or at least managed to the barest minimum.
- **Increased worker productivity:** No matter how strong the immune system is, there are times when it fails, especially when there is excess air pollution. As pollution is controlled, workers can now work for a longer period.
- **Helps improve indoor air quality:** Air pollution control helps to secure the quality of the air inside your house.

D-Challenges to fighting air pollution

Most of the world agrees that air pollution is a danger, but how do we conquer it? What's holding us back? Despite the global threat from air pollutants, it's hard for scientists to link them to a specific environmental disaster.

Developing countries contribute to a large share of pollution

Pollution of the environment is profitable for cooperation and for individuals. Simply put, this isn't their top priority. "We're asking very poor countries that are worried about where their next meal is coming from, or whether they can send their kids to school, to incur costs to reduce air pollution to benefit the world. And that's a tough ask for a policymaker inside of a developing country,"

Modern living is part of the problem

It's a tough pill to swallow, but modern conveniences like electricity, transportation, and air conditioning contribute to air pollution, and remedies potentially involve significant sacrifice and lifestyle change.

E-Main sources of air pollution

We will find the main sources of air pollution and divide our big problem into subprograms to make the analysis easier and clearer. After doing research we could divide the reasons for air pollution into two main sources.

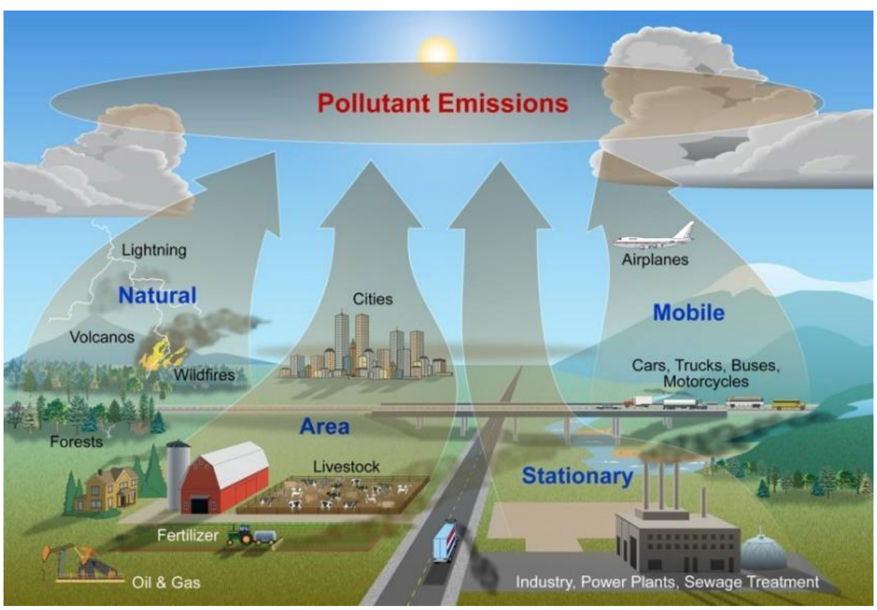

Figure 3 air pollutoin sources

1-Man-made (Artificial) sources

- **Stationary sources**
  - Fossil-fuel power plants and biomass power plants both have smoke stacks
  - Oil and gas sites that have methane leaks
  - Burning of traditional biomass such as wood, crop waste and dung.
  - Furnaces and other types of fuel-burning heating devices
  - Emission from Industries and Factories

- **Mobile sources**

Motor vehicles, trains ,marine vessels and aircraft as well as rockets and re-entry of components and debris. The air pollution externality of cars enters the air from the exhaust gas and car tires . Vehicles were reported to be "producing about one-third of all U.S. air pollution and are a major driver of climate change.

- **Area sources**
  - Agriculture and forest management strategies using controlled burns. Practices like slash-and-burn in forests like the Amazon cause large air pollution with the deforestation.
  - One of the most dangerous chemicals released during agricultural activity is ammonia. Insecticides, pesticides, and fertilizers pollute the atmosphere by releasing toxic chemicals.
  - In Europe, around 90% of ammonia emissions are due to agricultural pollution .

2-Natural sources

- Dirt from natural sources, typically big areas of land with little or no plant life.
- Methane, discharged by the breakdown of food (digestion) by animals, for example, cattle
- Smoke and CO from jungle fires.
- Volcanic activity, which emits sulfur, chlorine, and ash particulates
- Radon gas from radioactive decay within the Earth's crust.

Data Collection and Analysis

We collected data on the global level about:

•Co emissions from 1990 to 2020

• PM2.5 emissions from 1960 to 2017.

• Death rate due to air pollution from 1990 to 2019

1. CO emissions all over the years:

Breathing air with a high concentration of CO reduces the amount of oxygen that can be transported in the blood stream to critical organs like the heart and brain.

At very high levels, which are possible indoors or in other enclosed environments, CO can cause dizziness, confusion, unconsciousness, and death.

Very high levels of CO are not likely to occur outdoors. However, when CO levels are elevated outdoors, they can be of particular concern for people with some types of heart disease. These people already have a reduced ability for getting oxygenated blood to their hearts in situations where the heart needs more oxygen than usual. They are especially vulnerable to the effects of CO when exercising or under increased stress. In these situations, short-term exposure to elevated CO may result in reduced oxygen to the heart accompanied by chest pain also known as angina.

1.Numerically

Column1	Column2
Mean	4956.0582
Median	4603.5353
Standard deviation	1884.8824
Standard error	338.53487
Variance	3552781.7
Minimum	2666.008
Maximum	8366.9906
Range	5700.9826
Count	31

2.Graphically

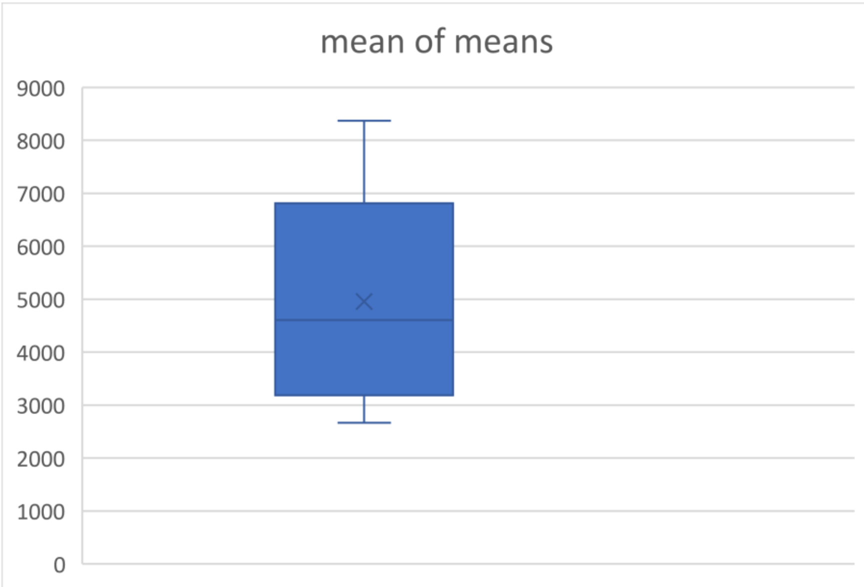

Figure 4 mean of means

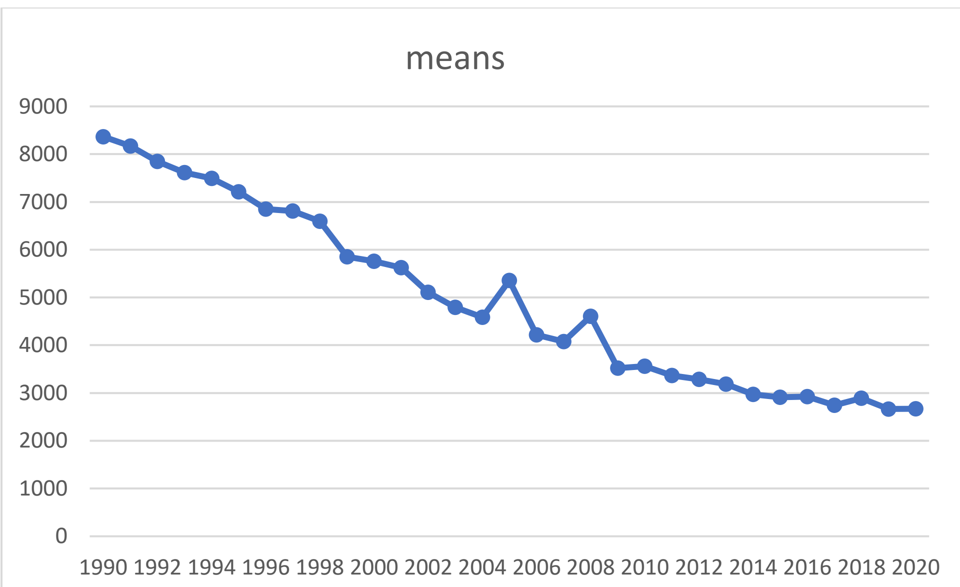

Figure 5 means

2.PM2.5 Emissions over the years:

Common resources:

sources of primary particles include fires, smokestacks, construction sites, and unpaved roads; sources of secondary particles include reactions between gaseous chemicals emitted by power plants and automobiles.

Environmental risks:

contributes to formation of haze as well as acid rain, which changes the ph balance of waterways and damages foliage, buildings, and monuments.

We calculated the mean of Data for each year and the following statistics is calculated for the mean of data samples.

1.Numerically

Mean	30.415877
Median	30.756269
Count	12
Maximum	31.342905
Minimum	28.768687
Range	2.5742186
Standard deviation	0.8457278
Standard error	0.2441406
Variance	0.7152555
Count	12

2.Graphically

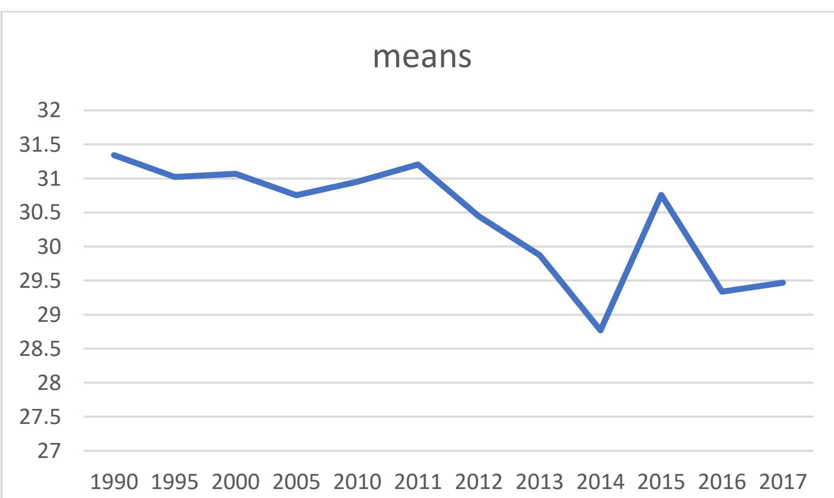

Figure 6 PM2.5 emission mean

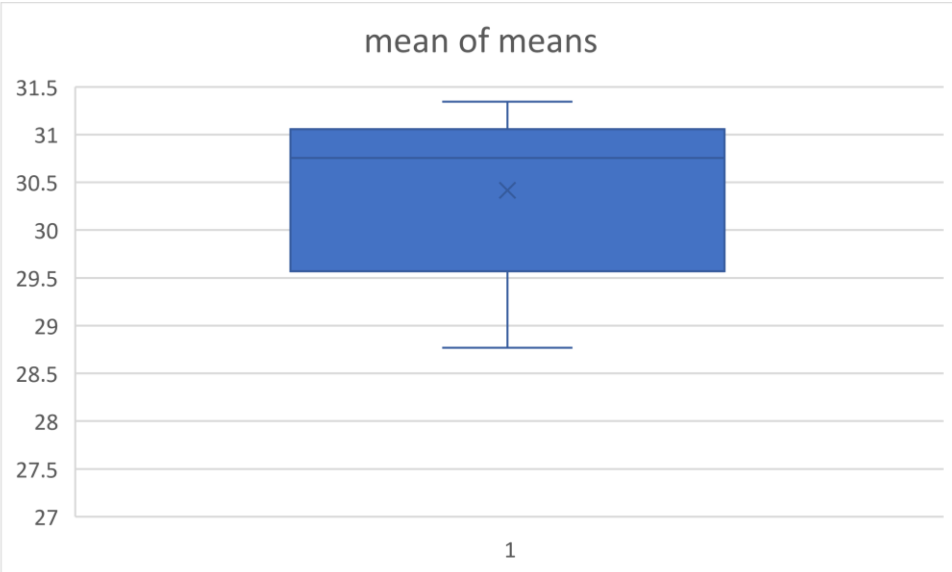

Figure 7 box plot for the means of pm2.5 emission

3.Death rates due to air pollution over the years

The World Health Organization (WHO) highlights air pollution as the number one reason for environment-related deaths in the world. It's estimated to be the cause of seven million premature deaths every year

At the global level, we see that since 1990, this percentage has been falling. In most countries, the likelihood of dying prematurely from air pollution has been declining over the last few decades.

Figure 9 total annual number of deaths by risk factor, measured across all age groups in Egypt, 2019(our world in data)

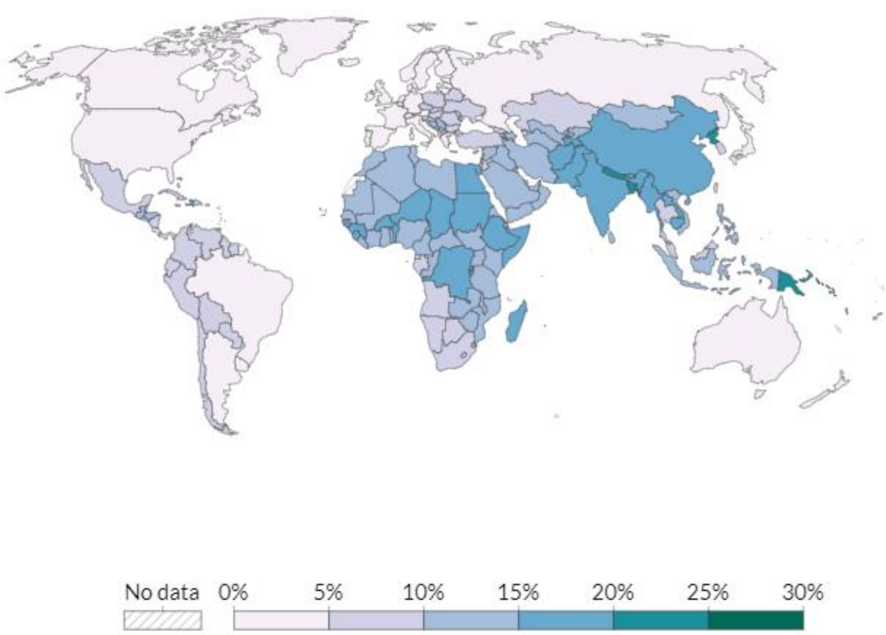

Figure 8 deaths from air pollution – from outdoor and indoor sources- Source: IHME, Global Burden of Disease

1.Numerically

Mean	70.179252
Standard deviation	17.221927
Var	296.59475
Sum	2105.3775
Count	30
Min	43.618798
Max	99.457837
Range	55.839038
Median	69.552332
Standard error	3.1442792

2.Graphically

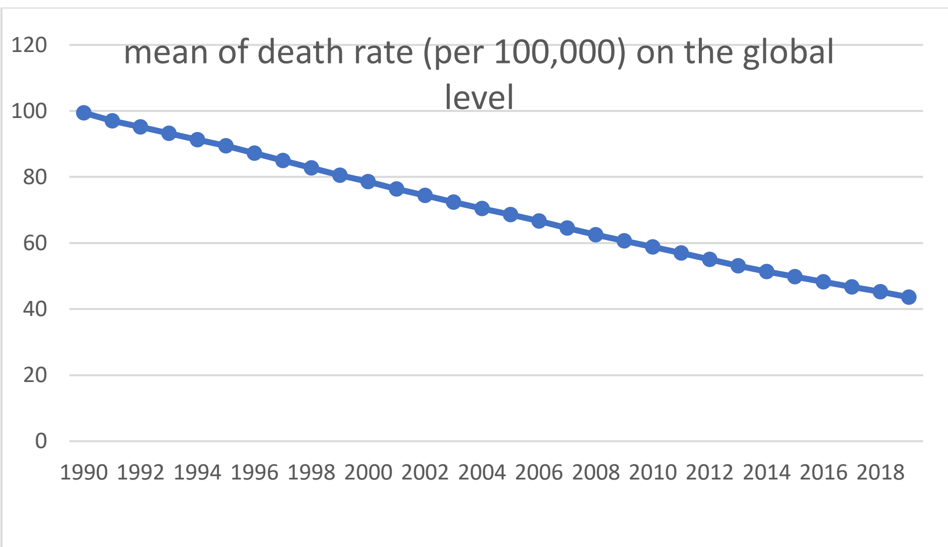

Figure 10 death rate mean

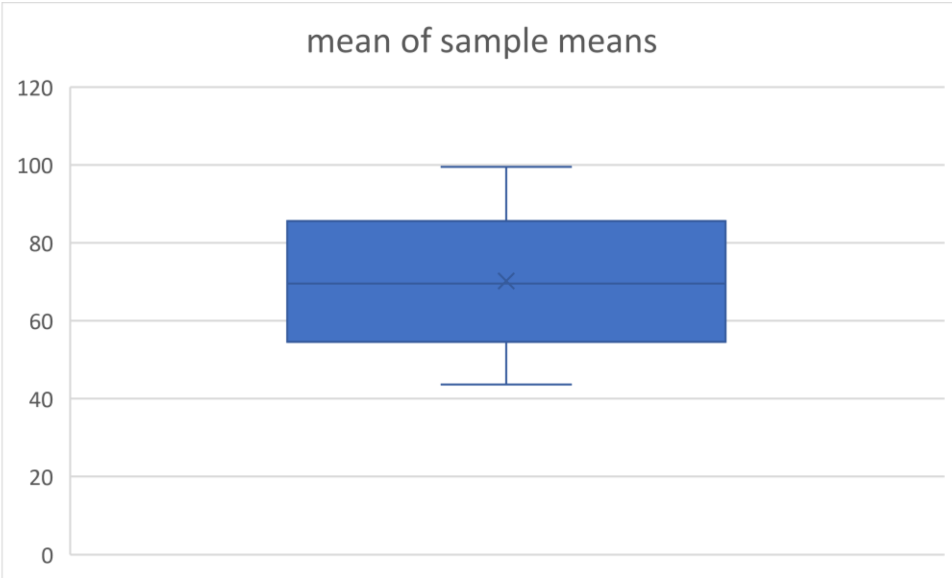

Figure 11box plot for death rate means

Methods of solutions

Hypothesis theorem

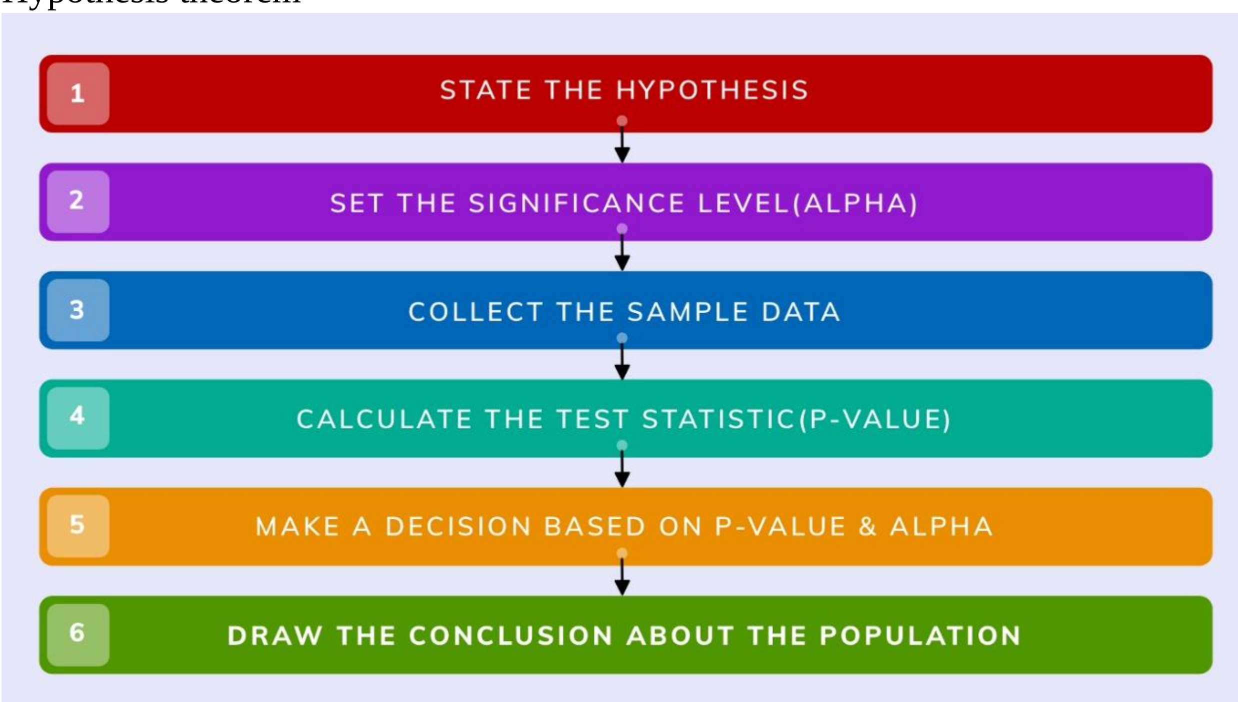

Figure 12 hypothesis theorem flow steps

Death rate due to air pollution from 1990 to 2019:

The data description showed that the death rate due to air pollution decreased by 60% from 1990 to 2019. How sure we are that this percentage, or this value will or won't decrease? Is there any possibility that death rate will increase? To know the answers to those questions we will use the hypothesis testing, Hypothesis testing is a systematic procedure for deciding whether the results of a research study support a particular theory which applies to a population. Hypothesis testing uses sample data to evaluate a hypothesis about a population, and there are some steps of hypothesis testing: The first is Specify the Null Hypothesis:

Null hypothesis (H0) is a statement that the death rate mean will equal 40 person per 100000 in 2023?

The second step is Specify the Alternative Hypothesis:

The alternative hypothesis (H1) is the statement that the death rate mean doesn't equal 40 per 100000

The third step is to Set the Significance Level a. The significance level (denoted by the Greek letter alpha— a) is generally set at 0.055. This means that there is a 5.5% chance that will accept the alternative hypothesis when the null hypothesis is true. The fourth step is Calculate the Test Statistic and Corresponding P-Value. The p-value describes the probability of obtaining a sample statistic as or more extreme by chance alone if the null hypothesis is true. This p-value is determined based on the result of the test statistic. And the step five is drawing a Conclusion and it is saying that P-value less than or equal significance level (a) greater than or equal Reject the null hypothesis in favor of the alternative hypothesis. Applying this algorithm on CO2 emissions data set using python code will give us those results

```
[1.0000e-02 5.1000e-01 9.2630e+01 6.0000e-02 1.4300e+00 3.6700e+00
 2.2952e+02 2.2240e+01 7.5000e+00 6.0880e+01 3.9000e+00 1.2000e-01
 1.0000e-02 1.0000e-02 6.4000e+00 1.0500e+00 1.2267e+02 1.5800e+00
 4.0200e+00 1.1100e+02 2.0000e-02 1.0252e+02 3.2000e-01 1.0000e-02
 1.1508e+02 1.4030e+01 4.5000e-01 1.7696e+02 3.0000e-02 4.3000e-01
 1.4500e+00 1.3375e+02 4.0000e-02 1.4373e+02 1.9700e+00 3.0000e-02
 1.3234e+02 2.7320e+01 9.9000e-01 2.2957e+02 1.1000e-01 1.0460e+01
 1.3696e+02 5.0000e-02 1.6663e+02 5.6100e+00 4.0000e-02 1.3828e+02
 3.9120e+01 1.6700e+00 2.4097e+02 6.7700e+00 6.4600e+00 1.3466e+02
 4.0000e-02 1.4981e+02 2.4400e+00 3.0000e-02 1.3235e+02 2.9290e+01
 1.0900e+00 2.3614e+02 1.3000e-01 3.5900e+00 0.0000e+00 1.0000e-02
 3.0000e-02 4.7000e-01 9.2000e-01 2.1380e+01 2.1773e+02 3.1000e-01
 4.0380e+01 1.0000e-02 6.9580e+01 4.6000e-01 6.0000e-02 1.0000e-01
 2.4400e+00 6.8900e+00 7.8850e+01 2.7746e+02 2.6300e+00 1.3427e+02
 1.0000e-01]
47.612117647058824
p-values 0.0046125588420384825
 we are rejecting null hypothesis
```

Figure 13 hypothesis testing for death rates results

Using z-interval with alpha = 0.055 we can find that Zcritical < Z so we tend to reject H0 and accept H1 that the true mean doesn’t equal 40 per 100000

By using interpolation, we found that in 2023 death rate mean will be 37.93, so the death rate is decreasing yearly with confidence 95.5%.

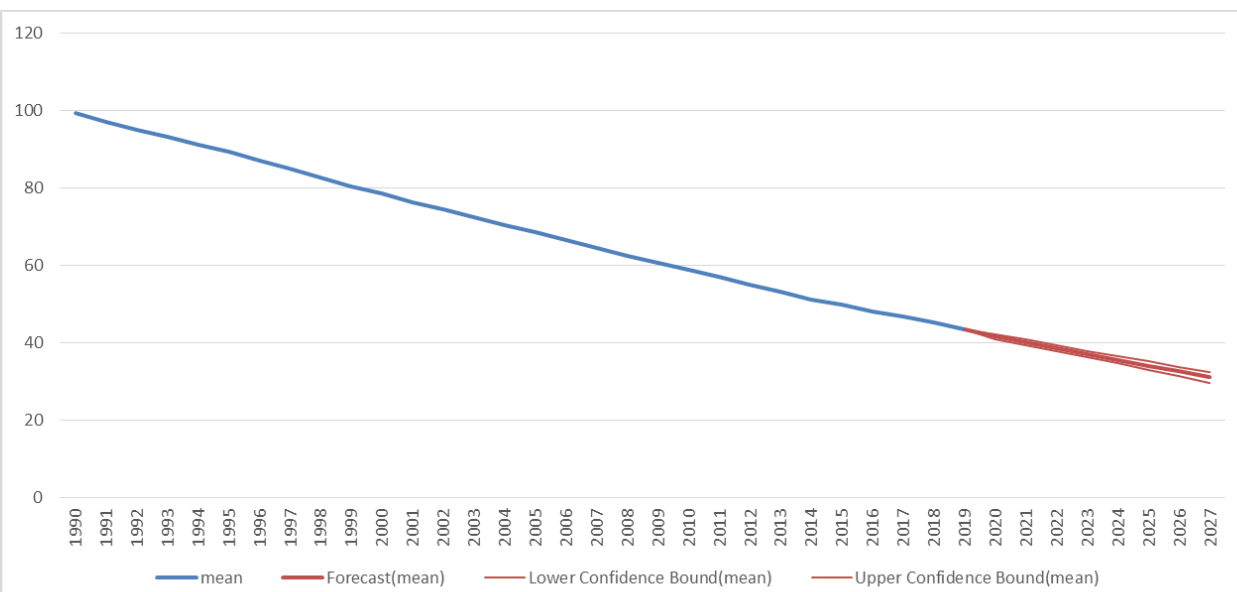

Figure 14death rate forecasting to 2027

PM2.5 Emissions over the years:

We noticed from data description of PM2.5 emissions dataset that the mean temperature in increasing over the years if we apply hypothesis testing as described in the previous algorithm on this dataset with significance level 1% and H0 is that true mean equals 30 (To assume the true population mean is degreasing) and H1 is the true population mean doesn't equal 32 we will get the following results

```
[58.06448  58.68747  60.36831  39.80279  42.58724  42.71914  18.08347
 24.94064  24.09334  13.46591  48.96254  40.71816  14.6804   52.02301
 51.17697  57.16109  44.34911  48.66031  40.06278  16.70844  16.66406
 16.69619  45.28857  51.65428  42.88246  62.74332  55.66797  53.85619
 54.10316  52.63831   7.311075 16.6998   16.67343  16.68744  44.93633
 51.29116  42.60216  62.29024  55.80675  54.37185  54.62999  52.25184
  7.321048 53.89228  53.06037  59.35402  46.04589  50.52036  44.59379
 17.75361  17.69023  17.74335  47.03639  53.57572  43.85541  63.1777
 57.01834  54.46832  54.71259  54.63514  21.8724   26.14209  18.07617
 11.08953  39.80006  16.53495  13.04341  48.86025  15.39804  26.77918
 14.12448  19.5811   23.90916  24.68348  17.55445  10.56968  32.30682
 40.21579  15.55618  25.44909  17.4324   21.83986   8.430091 81.29063
 18.05281  38.07174  63.71657  10.40503 ]
36.84434288636363
p-values 0.014705689868863774
 we are rejecting null hypothesis
```

Figure 15 PM2.5 hypothesis testing results

Using z-interval with alpha = 0.01 we can find that Zcritical < Z so we tend to reject H0 and accept H1 that the true mean doesn't equal 32 But how can we make sure this analysis is right??????

By doing interpolation in the PM2.5 emissions dataset we get that by 2024 the emission mean will be 26.5666.

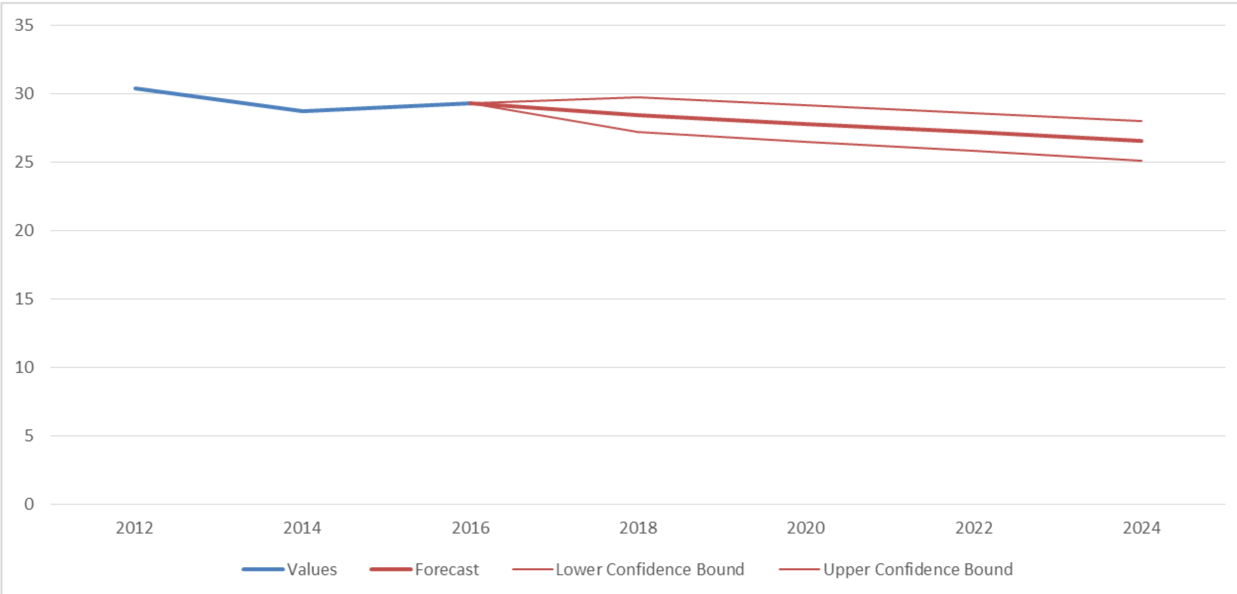

Figure 16PM2.5 emission prediction to 2024

CO emissions over the years:

As the data description illustrates that the CO emission has been decreasing from 1990 to 2020. If we apply hypothesis testing as described in the previous algorithm on this dataset with significance level 5% and H0 is that true mean equals 2300 (To assume the true population mean is decreasing) and H1 is the true population mean doesn't equal 2300 we will get the following results

```
[  451.33   657.54   416.13  2202.67   260.87   278.66    87.11  1433.92
   287.32   151.98  1907.27  1244.78   339.     870.92   392.63   157.72
  1099.25   817.1   2023.31  8388.3    439.07  4199.09  3747.88   762.26
   502.42   118.48   247.37   212.83 11325.1  11596.59  3510.26  2758.48
   696.13   203.82   174.09    33.06   537.29   364.98  5187.52  2155.
  1971.24   534.76   851.5   6783.71  3962.82  6351.94  3927.13   130.07
  7280.03  4168.71  2641.25  3068.67   537.76   463.94   118.73   177.68
   135.96  2266.32  4750.21 10562.88 10038.81  9373.53  9186.98  8650.13
  8034.19  7764.68  7277.93  6607.76  6208.84  5995.77  5689.93  5751.19
  5244.16  4655.09  3764.    1478.68   577.34   256.15   631.72  8034.19
  6312.69  1064.82   902.39    59.82   377.  ]
3021.913294117647
p-values 0.04441515371022251
 we are rejecting null hypothesis
```

Figure 17 Co emission hypothesis results

Using z-interval with alpha = 0.05 we can find that Zcritical < Z so we tend to reject H0 and accept H1 that the true mean doesn’t equal 2300

By using interpolation, we found that in 2023 CO emission mean will be 3084.38, so the CO emission decreasing with confidence 95%.

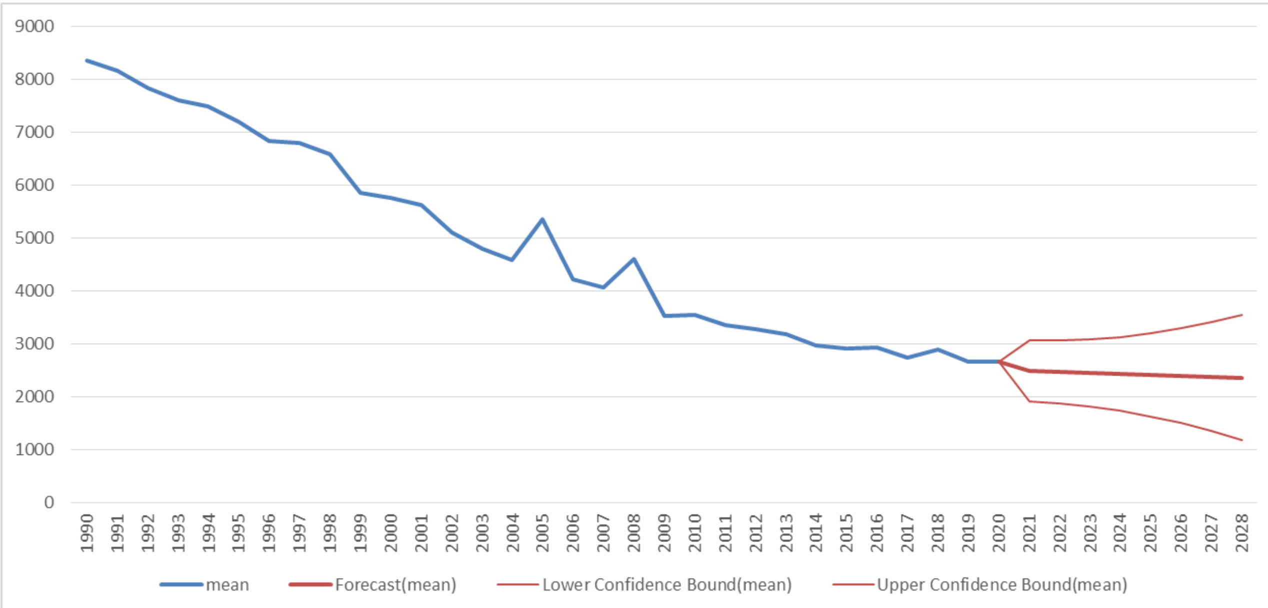

Figure 18 CO emission forecasting to 2028

Linear Regression modeling

Co Vs death rated:
Using the mean dataset of death rated for each value of Co emission. We found the following results

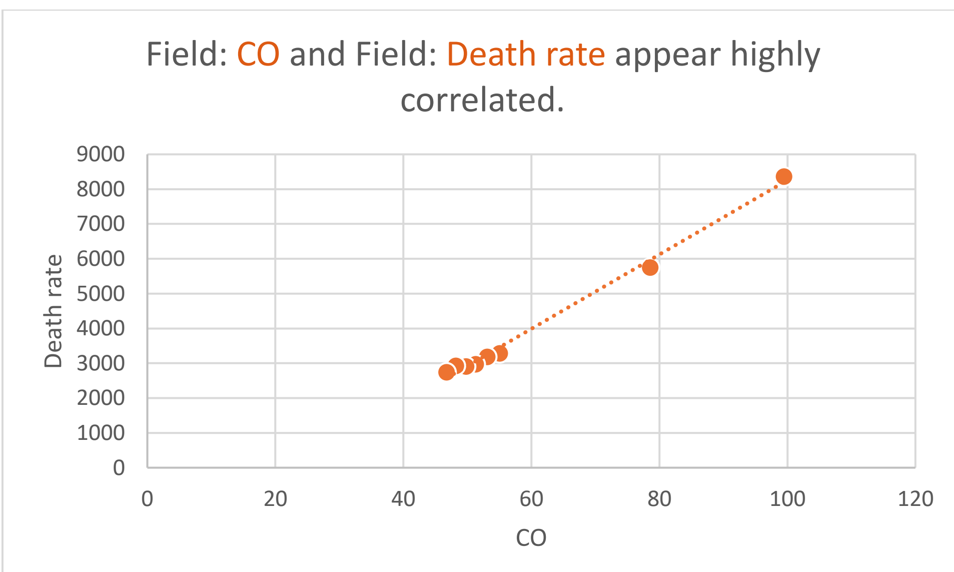

Figure 19 regression graph between deathrates and CO emission

And applying python code to find the regression equation we got:

```
coefficient of determination: 0.9934483496402665
intercept: -2414.886601163331
slope: [106.72506832]
```

Figure 20 y intercept and slop

$$Death\ rate = 106.725 * CO\ emission - 2414.8866$$

We did some prediction for the Death rates depending on the regression equation:

CO emission	Death rate (per 100000)
100	8257.62023085
120	10392.12159726
90	7190.36954765

PM2.5 Vs death rate:

Using the mean dataset of death rated for each value of PM2.5 emission. We found the following results

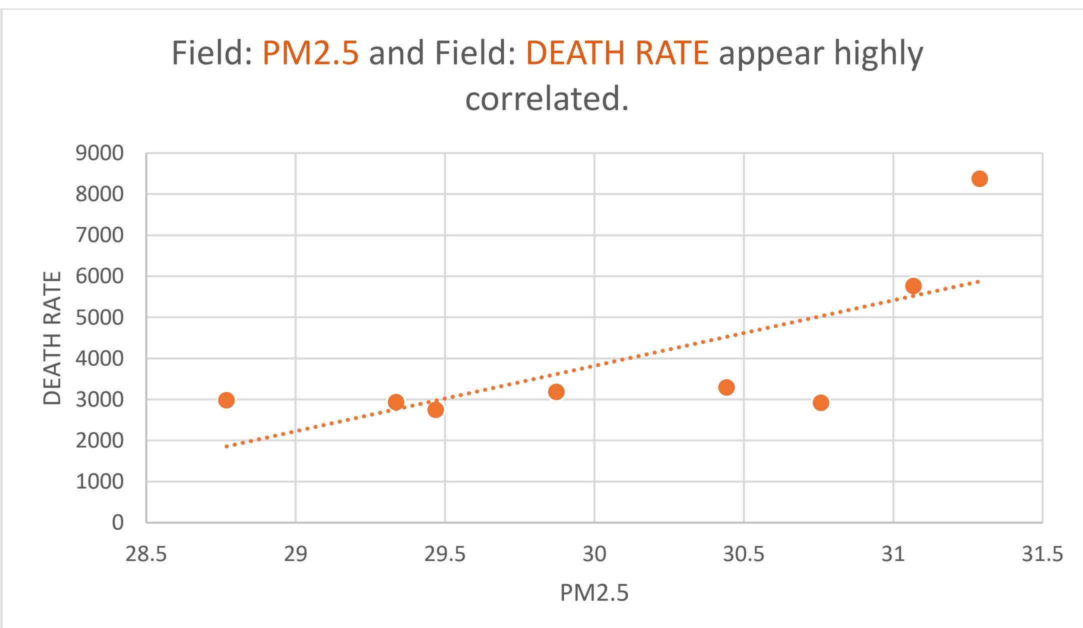

Figure 21 PM2.5 Vs death rate regression graph

And applying python code to find the regression equation we got:

```
coefficient of determination: 0.5123079423517286
intercept: -43998.248173005406
slope: [1593.91152577]
```

Figure 22 slope and y intercept

$$Death\ rate = 1593.9 * CO\ emission - 43998.25$$

We did some prediction for the Death rates depending on the regression equation:

PM2.5 emission	Death rate
30	3819.09760009
40	19758.21285779
50	35697.32811549
60	51636.44337319

So now we can interpret that there is a strong dependence between death rates with CO and PM2.5 emissions.

Analysis of Results

From the data above we concluded that air pollution has been declining in the past few years mainly because of the lockdown due to covid 19 and the deaths due to air pollution also has been decreasing. Despite great progress in air quality improvement, approximately 102 million people nationwide lived in counties with pollution levels above the primary NAAQS in 2021.

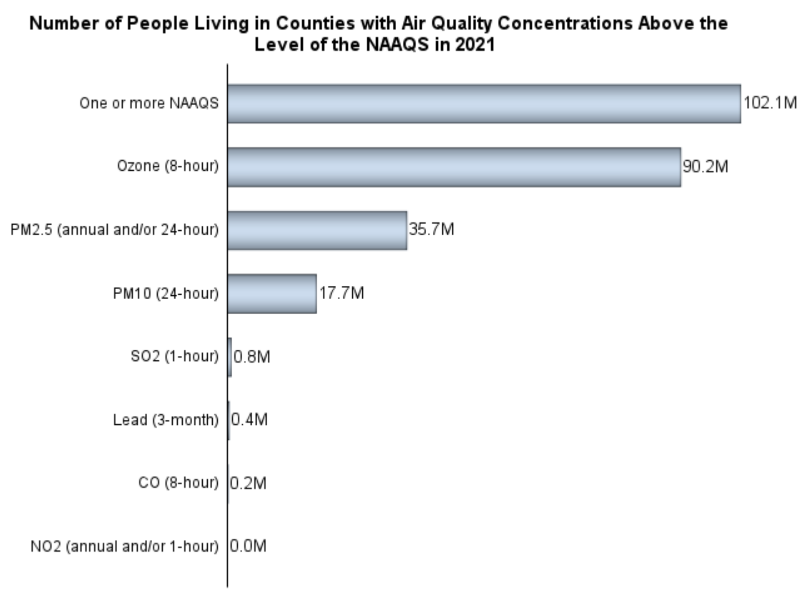

Figure 23 air quatlity and living concentration

In addition, from 1990 to 2017 emissions of air toxics declined by 74 percent, largely driven by federal and state implementation of stationary and mobile source regulations.

In recent years, EPA has acted to dramatically improve air quality by designing and developing programs that, when fully implemented, will achieve significant reductions in air emissions. The associated air quality benefits will lead to improved health, longevity, and quality of life for all people over the world.

Now let's consider maintaining the decrease of death rates and keeping our planet healthy:

The choices that we make every day can help to make a difference. Here are some simple steps you can take:

- Commute smart by walking or riding to work or the shops instead of driving. Motor vehicle emissions remain the most significant source of most common air pollutants.
- Choose a fuel-efficient vehicle next time you are replacing your car. The Green Vehicle Guide provides ratings on the environmental performance of new vehicles sold in Australia.
- Save energy, by turning off the television and make sure you flick the light switch when you leave the room. Not only will you save money on your electricity bill, you will be reducing emissions from coal-fired electricity plants.
- Buy energy -efficient appliances. Check the energy rating label when buying new or second-hand electrical appliances. More stars mean less emissions. The Energy Rating web site will help you select energy efficient appliances.
- Use environmentally safe paints and cleaning products whenever possible.

Conclusion and Expectations

In conclusion, air is essential for life, so we must limit air pollution due to its hazard effect on all forms of life. Data description and applying hypothesis testing illustrates that the air pollution percentage decreased in the past few years which leads to decreasing the annual deaths due to air pollution. As we have seen in the previous analysis, a lot of people live in areas with high concentration of air pollutants, so we have to work on improving air quality and spreading awareness to fight for the next generations and sustain life on earth.

We expect with more awareness spreading, people will actively react to this crisis, stop using polluting materials and move to more organic, eco-friendly products in their everyday lifestyle.

References

- ✓ Https://www.nps.gov/subjects/air/sources.htm
- ✓ Https://byjus.com/chemistry/air-pollution/
- ✓ Air Pollution - Causes, Sources, Types, Consequences, Prevention and FAQ (vedantu.com)
- ✓ Air Pollution: Causes, Sources, Types, Consequences, Control (collegedunia.com)
- ✓ How is air pollution caused? What are its effects on our health and the environment? | clientearth
- ✓ Air pollution - Ozone | Britannica
- ✓ Https://scied.ucar.edu/learning-zone/air-quality/effects-air-pollution
- ✓ Https://www.airbreath-oxy.com/en/why-is-air-important/
- ✓ Https://ourworldindata.org/air-pollution
- ✓ Https://bettergoods.org/air-pollution-statistics/

Python code:

```
from scipy.stats import ttest_1samp
import numpy as np
ages = np.genfromtxt("co2.csv")
print(ages)
ages_mean = np.mean(ages)
print(ages_mean)
tset, pval = ttest_1samp(ages, 40)
print("p-values",pval)
if pval < 0.005:    # alpha value is 0.005 or .5%
    print(" we are rejecting null hypothesis")
else:
    print("we are accepting null hypothesis")
```

Figure 24 hypothesis theorem code

```
import numpy as np
from sklearn.linear_model import LinearRegression
emission= np.array([31.28930051,31.06815112,30.44238501,29.87204398,28.7686
Death_rate = np.array([8366.9906,5756.015,3286.776,3183.189,2974.443,2911.8
model = LinearRegression()
model.fit(emission, Death_rate)
r_sq = model.score(emission, Death_rate)
print(f"coefficient of determination: {r_sq}")
print(f"intercept: {model.intercept_}")
print(f"slope: {model.coef_}")
new_model = LinearRegression().fit(emission, Death_rate.reshape((-1, 1)))
emission=np.array([30,40,50,60]).reshape((-1, 1))
Death_rate_pred = model.predict(emission)
print(f"predicted response:\n{Death_rate_pred}")
Death_rate_pred = model.intercept_ + model.coef_ * emission
```

Figure 25 linear regression model code

Table of figures