AH Provisional Diabetes Death Counts, 2020

Prepared By:

Michael Campos

Prepared For:

California State University of Los Angeles

Healthcare Data Analytics CIS – 4210

Professor Balan

Objective of the study:

When looking at the leading cause of death and disability in the united states, it is chronic diseases. According to the Centers for Disease Control and Prevention (CDC), chronic disease is defined as "conditions that last one year or more and require ongoing medical attention or limit activities of daily living or both". These conditions are heart disease, cancer, chronic lung disease, stroke, Alzheimer's disease, diabetes, chronic kidney disease. Another fact about the chronic disease from CDC is that six in ten adults have one chronic disease, and four in ten adults have two or more. While chronic diseases are not curable, they can be treated to be controlled or stable.

With COVID-19 being introduced in 2020 and affecting many individuals, it strongly affected individuals with one or more chronic diseases. Before discussing how there is a correlation between COVID-19 and chronic disease, I first need to explain what covid-19 means. According to Cleveland Clinic, Coronavirus disease (COVID-19) is defined as "a family of viruses that can cause respiratory illness in humans". Now how does COVID-19 strongly affect chronic diseases? Well, for example, diabetes causes high blood pressure, and due to this, it causes damage to the body's blood vessels. When there is damage to the blood vessels, it compromises individual lungs. Because of COVD-19, causing respiratory illness will lead the individual to more complications, leading to death.

Throughout this paper, you will see a data description describing what the data set is all about. For example, what each column means and what the rows contain. There will also be three analysis questions:

- Which age group has the highest death count having COVID-19 and hypertensive disease while also showing whether it is most common in males or females?
- What month has the highest death count for diabetes as an underlying or contributing cause of death?
- What is the death count for having COVID-19 with Obesity for each age group and sex?

For each question of analysis, there will be visualizations to illustrate the relationship of data better. Then a statistical summary will be presented to summarize the observations, trends, patterns made. Finally, a statistical test will be performed, such as one-way frequency, correlation analysis, and t-test.

In conclusion, chronic disease is the leading cause of death and disability in the united states. Having both COVID-19 and chronic disease can lead individuals to have complications in their health, leading to death. With the data set "AH Provisional Diabetes Death Counts, 2020," a series of analyses are conducted to visualize the death counts of chronic disease and COVID-19.

Data Set URL: AH Provisional Diabetes Death Counts, 2020 - CKAN (data.gov)

Data Set Description:

Data Set Name	WORK.IMPORT	Observations	226
Member Type	DATA	Variables	16
Engine	V9	Indexes	0
Created	05/12/2021 20:38:07	Observation Length	136
Last Modified	05/12/2021 20:38:07	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
4	AgeGroup	Char	11	$11.	$11.
15	C19PlusChronicLiverDiseaseAndCir	Num	8	BEST12.	BEST32.
13	C19PlusChronicLowerRespiratoryDi	Num	8	BEST12.	BEST32.
9	C19PlusDiabetes	Num	8	BEST12.	BEST32.
10	C19PlusHypertensiveDiseases	Num	8	BEST12.	BEST32.
12	C19PlusHypertensiveDiseasesAndMC	Num	8	BEST12.	BEST32.
14	C19PlusKidneyDisease	Num	8	BEST12.	BEST32.
11	C19PlusMajorCardiovascularDiseas	Num	8	BEST12.	BEST32.
16	C19PlusObesity	Num	8	BEST12.	BEST32.
6	COVID19	Num	8	BEST12.	BEST32.
1	Data as of	Num	8	MMDDYY10.	MMDDYY10.
3	Date_Of_Death_Month	Num	8	BEST12.	BEST32.
2	Date_Of_Death_Year	Num	8	BEST12.	BEST32.
8	Diabetes.mc	Num	8	BEST12.	BEST32.
7	Diabetes.uc	Num	8	BEST12.	BEST32.
5	Sex	Char	10	$10.	$10.

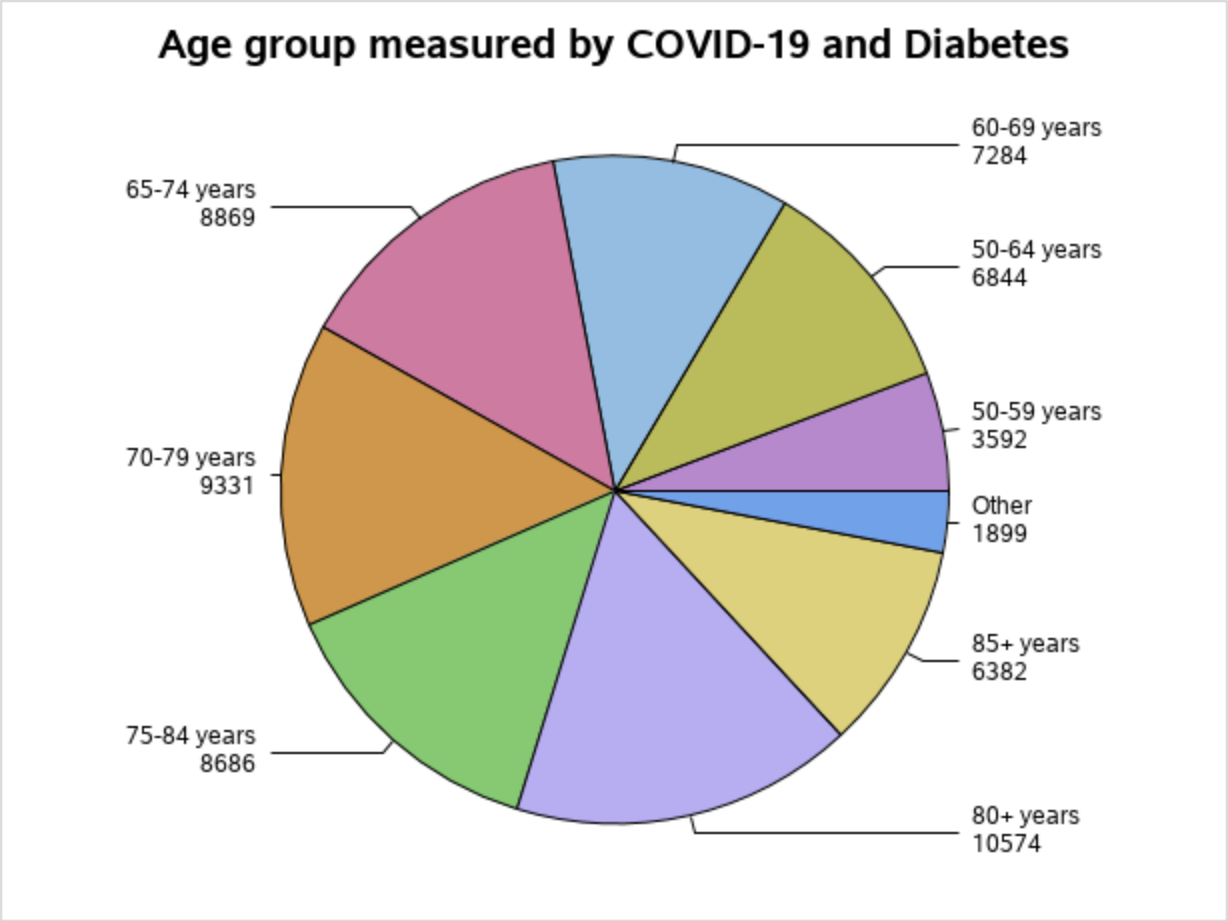
The data set "AH Provisional Diabetes Death Counts, 2020" provides information about causes of death, sex, age, and the list of months. The data set has 16 columns, and 227 rows, the data types used in this data set are numeric and varchar. When looking at the names of the 16 columns, they are date as of, date of the year, age group, sex, covid19, diabetes.mc, diabetes.uc, c19plus diabetes hypertensive diseases, c19plus major cardiovascular diseases, c19plus hypertensive diseases and mcvd, c19plus chronic lower respiratory disease, c19plus kidney disease, c19plus chronic liver disease and cirrhosis, and c19plus obesity. Before describing what

each column signifies, c19plus is described as having COVID-19 as underlying or contributing cause of death. Another thing to notice is that diabetes.mc is diabetes as underlying or contributing cause of death and diabetes.uc diabetes as underlying cause of death.

The first column, "data as of," is tell us when it has last been updated, which is 10/20/2020. Due to this, all the rows have 10/20/2020 meaning the data set is all up to date. The second column," date of death year," implies when the individual has passed away. In this case, the data set is focusing on the year 2020. This means that the second column rows consist of 2020. The third column, "date of death month" consists of 9 months. This column implies what month was the death of the individual from January to October. The fourth column, "age group," is grouped by ages. These ages are 18 and below, 18 - 29 years, 30 - 39, 40 - 49 years, 50 - 59 years, 50 - 64 years, 60 - 69 years, 65 - 74 years, 70 -79 years, 75 - 84 years, 80+ years, 85+ years, unknown age. The fifth column, " sex," consists of males and females. This will let us know whether the individual who died was male or female. Finally, from column 7 to 16 it provides the provisional death counts of each disease.

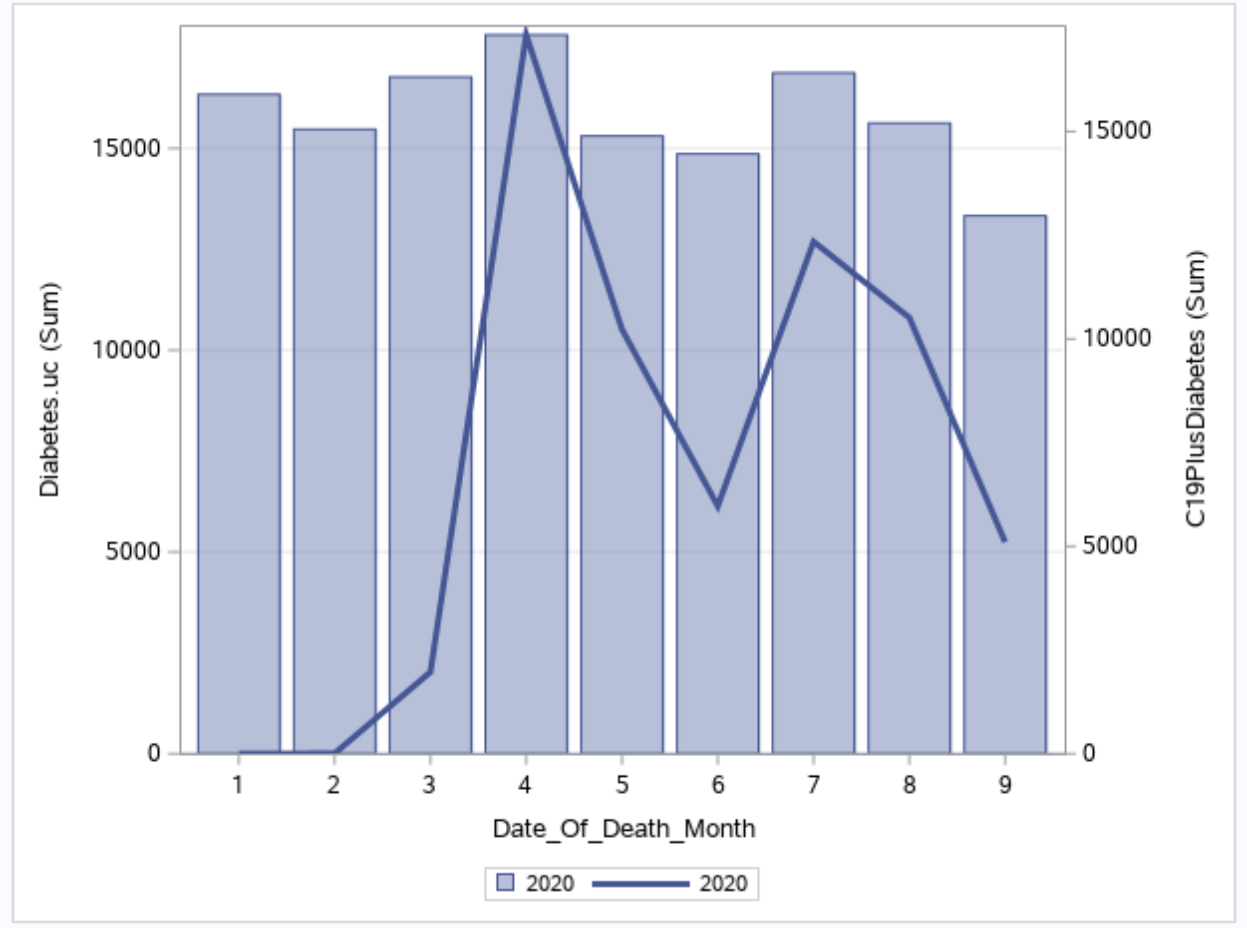
Question of Analysis:

What is the correlation between age group and hypertensive disease?



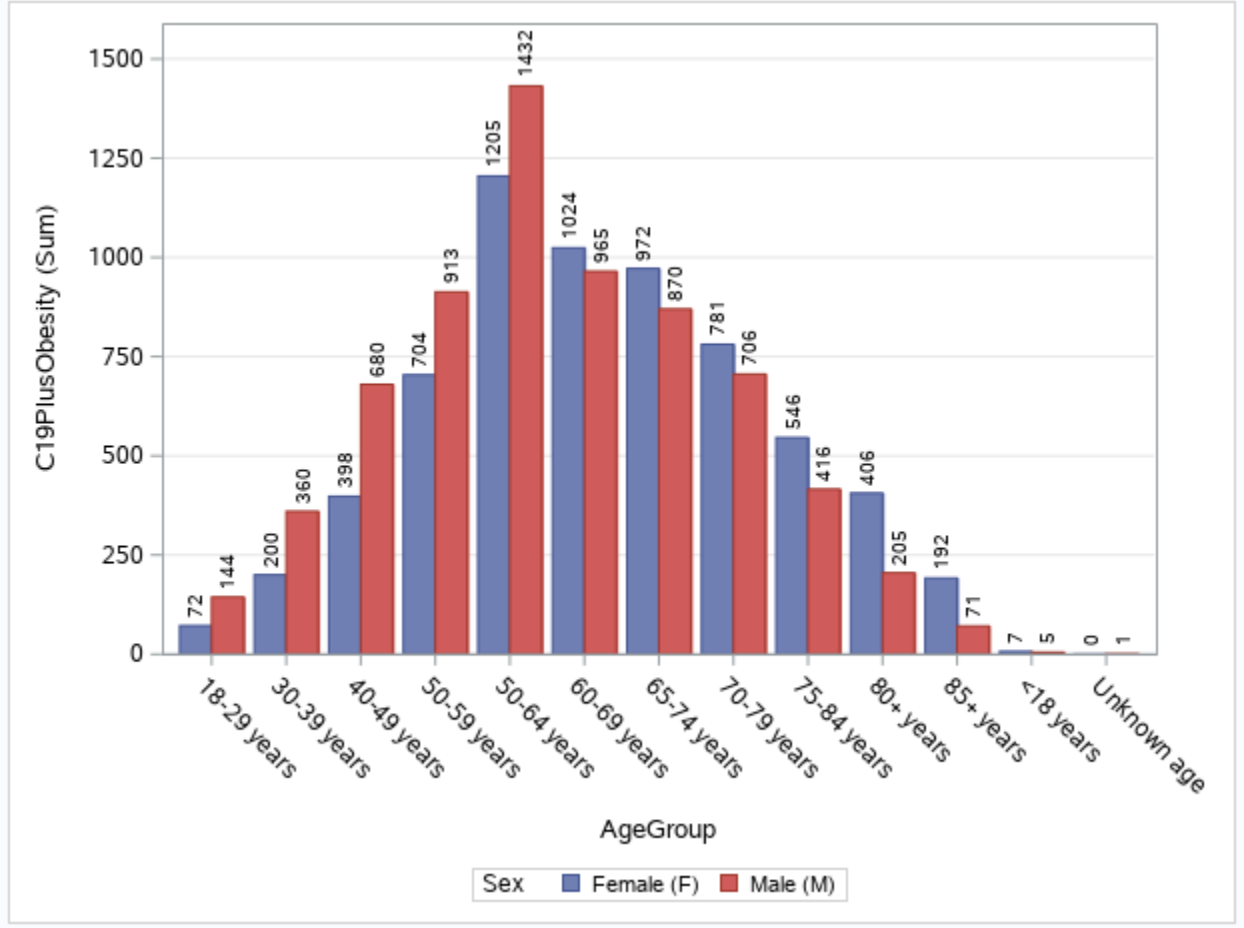
The chart above shows the different categories of age group measured by death counts by individuals who have COVID-19 and diabetes. One can see that the highest death count in the age group category is 80+ years old, and the lowest death count is 50-59 years old. According to National Institute on Aging (NIH), diabetes is commonly found in older adults due to having poor health. As you can see in the pie chart individuals, start to die at 50 years old. This means that to tackle this issue, individuals diagnosed with pre-diabetes need to start treating it early on to not lead to severe complications in the future.

What month has the highest death count for diabetes as the only contributor or having both COVID-19 and Diabetes?



The bar-line chart above shows the death count of diabetes as an underlying cause of death and COVID-19 with diabetes as an underlying cause of death from the first nine months of the year 2020. As you can see, diabetes as an underlying death is somewhat symmetric, meaning that it has been an issue throughout the year. However, when looking at COVID-19 with diabetes as the underlying cause of death, there have been two spikes throughout the year. This also correlates with diabetes as an underlying cause of death on having a huge amount of death counts in April. While there seem to be fewer deaths in the second spike in the second spike by having COVID-19 with diabetes, there is still a great amount of death count for diabetes as the underlying cause.

What is the death count for having COVID-19 with Obesity for each age group and sex?



The bar chart above shows the death count of COVID-19 with obesity for each age group sex. The purpose of this analysis is to show how COVID-19 affected many individuals to being obese due to being quarantine. As you can see, it greatly affected ages 40-79 from having 680 - 1432 death counts. Another pattern seen in the bar chart is that from ages 18-64, males have a higher death count. While from ages 60-85, females have a higher death count.

Statistical Summary

Analysis Variable : Diabetes.mc						
Mean	Std Dev	Minimum	Maximum	Median	N	Mode
2306.33	1939.94	0	7200.00	2437.50	226	6.0000000

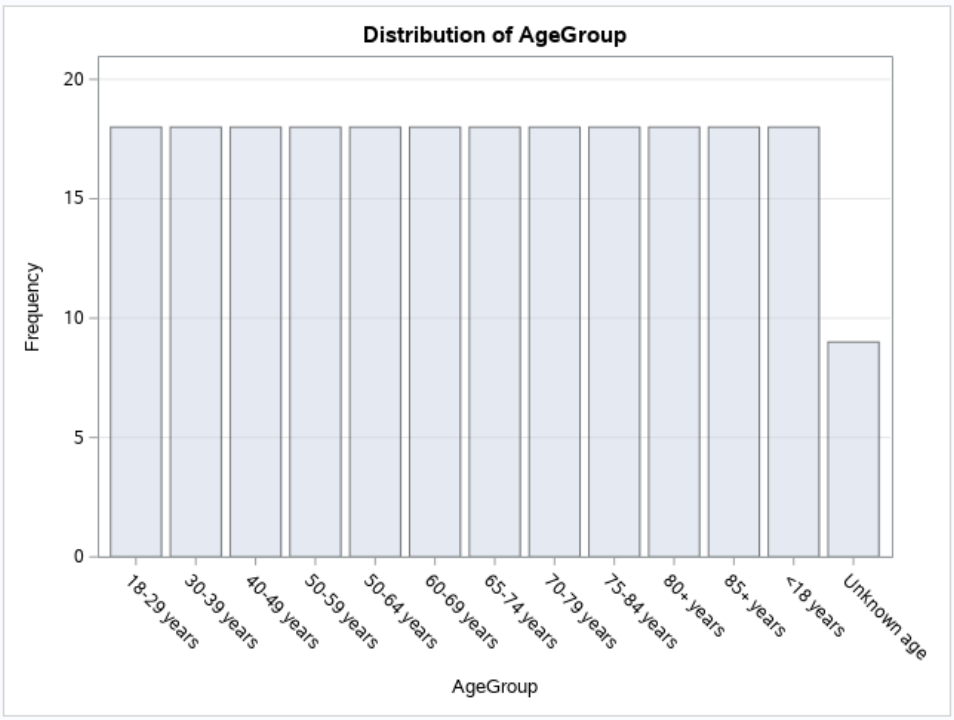
The best way to provide a brief description of data is through summary statistics. The table above describes the statistical averages of the column diabetes.mc, which is diabetes as an underlying or contributing cause of death. When looking at the table, the mean is 2306.33, meaning that is the average death count of diabetes. The median for diabetes death count is 2437.50, meaning that the value 2437.50 appears in the middle. The mode for the table is 6.0, meaning that are 6.0 values that appear more often in the table. The maximum of death counts is 7200.00, and the minimum is 0. This shows the highest and lowest death count for diabetes. N is signified as the number of observations, and in this case, there were 226 observations made. Lastly, the standard deviation is 1939.94 indicates that how far the value is. In this case, the standard deviation is close to the mean, indicating a low standard deviation.

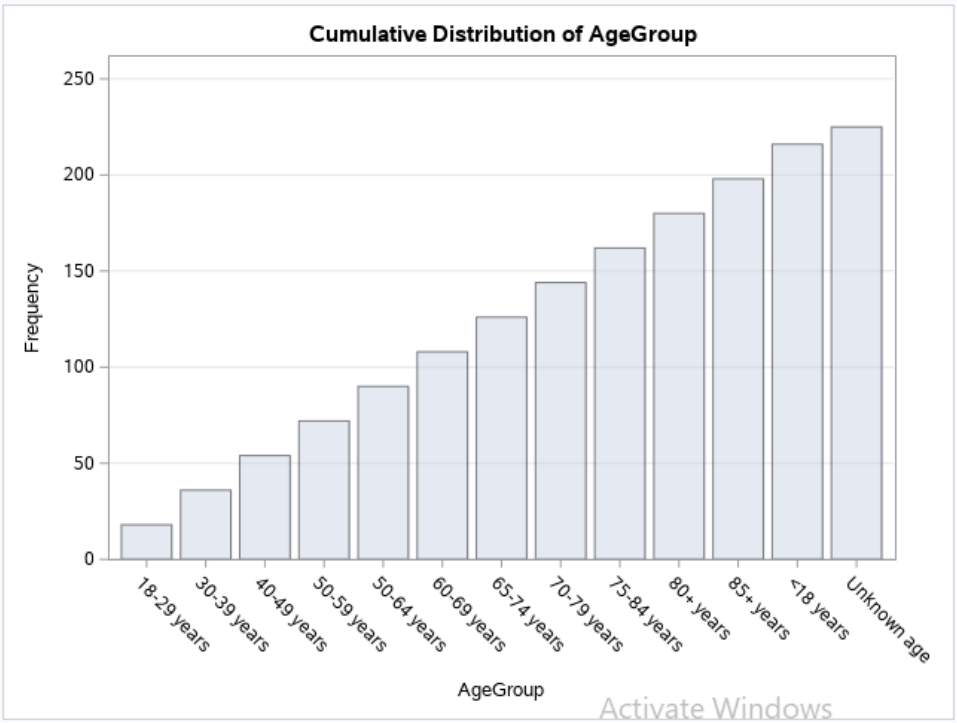
Statistical Tests

One-way frequency

The FREQ Procedure

AgeGroup	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18-29 years	18	8.00	18	8.00
30-39 years	18	8.00	36	16.00
40-49 years	18	8.00	54	24.00
50-59 years	18	8.00	72	32.00
50-64 years	18	8.00	90	40.00
60-69 years	18	8.00	108	48.00
65-74 years	18	8.00	126	56.00
70-79 years	18	8.00	144	64.00
75-84 years	18	8.00	162	72.00
80+ years	18	8.00	180	80.00
85+ years	18	8.00	198	88.00
<18 years	18	8.00	216	96.00
Unknown age	9	4.00	225	100.00
Frequency Missing = 1				





One-way frequency generates frequency based on the data set. In this case, the table above called "The FREQ Procedure" is a one-way frequency based on the age group column from the data set

"AH Provisional Diabetes Death Counts, 2020". The column age group provides information about the age and it group by roughly ten years apart, for example, 18-29 years, 30-39 years, etc. For each age group, there is a frequency of 18 except "unknown age," which has a frequency of 9. The way frequency is defined is by looking at the number of observations made for each age group. The percentage column shows the percentage of the total amount of observations made in frequency. The cumulative frequency is made up by adding frequency that comes before the frequency you are working with. For example, the age group 18-29 has a frequency of 18; this means that the cumulative frequency is 18 due to not having a frequency before it. In the age group, 30-39, it has a frequency of 18; the cumulative frequency would be 36 because the frequency before this one was 18, so by adding 18+18, you get 36. The cumulative percent column works the same way as the cumulative frequency column; instead of adding the frequency, you add the percent.

Correlation Analysis

1 With Variables:	COVID19
1 Variables:	Diabetes.mc

Pearson Correlation Coefficients, N = 226	
	Diabetes.mc
COVID19	0.70421

Correlation analysis provides statistics associations among two variables. In this case, the table above is showing the relation diabetes has with COVID-19. As you can see, there is a correlation of 0.70421, which means a 70 percent correlation between them. The reason being is that diabetes.mc and COVID-19 both have to do with underlying or contributing cause of death.

T-Test

Variable: Diabetes.mc
Sex = Female (F)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.906183	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.165327	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.500773	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.446542	Pr > A-Sq	<0.0050

Variable: Diabetes.mc
Sex = Male (M)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.885159	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.169132	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.729554	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.822664	Pr > A-Sq	<0.0050

Variable: Diabetes.mc

Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female (F)		112	2060.0	1807.8	170.8	0	7200.0
Male (M)		114	2548.3	2040.5	191.1	0	6590.0
Diff (1-2)	Pooled		-488.3	1928.7	256.6		
Diff (1-2)	Satterthwaite		-488.3		256.3		

Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female (F)		2060.0	1721.5	2398.5	1807.8	1598.1	2081.5
Male (M)		2548.3	2169.7	2926.9	2040.5	1805.7	2346.2
Diff (1-2)	Pooled	-488.3	-994.0	17.3898	1928.7	1765.5	2125.5
Diff (1-2)	Satterthwaite	-488.3	-993.4	16.8774			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	224	-1.90	0.0583
Satterthwaite	Unequal	221.66	-1.90	0.0581

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	113	111	1.27	0.2020

A T-test is used to determine an important difference between the one, two, or paired variables. In this case, we are using a two-sample test, meaning two variables will be used. The two variables we are examining are diabetes.mc, and sex. The result shows that Pr > |t| is 0.05 for both pooled and Satterthwaite. Due to this, it means that there is a statical significance between the two groups for the null hypothesis and alternative hypothesis because it equals the p-value of 0.05.

Work Cited

"About Chronic Diseases." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 28 Apr. 2021, www.cdc.gov/chronicdisease/about/index.htm.

"Chronic Diseases in America." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 12 Jan. 2021, www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm.

"Diabetes in Older People." *National Institute on Aging*, U.S. Department of Health and Human Services, www.nia.nih.gov/health/diabetes-older-people.

Nania, Rachel. "Chronic Conditions and Increased COVID-19 Risks." *AARP*, 1 Apr. 2020, www.aarp.org/health/conditions-treatments/info-2020/chronic-conditions-coronavirus.html.

Team, Healthy Aging. *The National Council on Aging*, 23 Apr. 2021, www.ncoa.org/article/the-top-10-most-common-chronic-conditions-in-older-adults.