Adiposity Statistical Consulting Project

Introduction

of adults . Even more shocking, in the United States alone, 42.4% of are obese (CDC) a body-mass index (BMI) of 30 or greater. In order to calculate someone's BMI the person's weight in kilograms is divided by their height in meters squared. Obesity can cause severe conditions heart disease, stroke, type 2 diabetes, and even cancer (CDC).

According to the Centers for Disease Control & Prevention (CDC) in the US 15 million plus people are affected by Chronic Obstructive Pulmonary Disease (CDC). In the United States alone, COPD is the cause of 150,000 deaths (CDC). very minutes an American dies from COPD (CDC). Currently, COPD does not have a cure, only treatment plans that can help patients have a better quality of life.

here are different subtypes of COPD have unique distributions and differential risk of mortality (Young). There are three different groups in which the high-risk COPD participants fall into: airway-predominant disease only (APD-only), emphysema-predominant disease only (EPD-only), and combined APD-EPD (Young). The Body mass index-Obstruction-Dyspnea-Exercise-capacity (BODE) index is specially made to identify individuals with COPD who are at an increased risk of mortality (Young).

The purpose of this project was to utilize data COPDGene to see if there is a correlation between obesity Airway Disease COPD. We hypothesize that due to inflammation, Airway Disease metabolic disease (more obesity/fat, cardiovascular disease, diabetes, etc). ata was collected by COPDgene investigators from the University of Iowa. This was an observational study 73 different pieces of information were measured attempt to affect the outcome.

Methods

Variables: There were 73 different pieces of information obtained from the willing participants in this project. We will be focusing on the following 12 variables in this project:

- Age (phase 1)
- Gender
- Race
- Cigarette Smoking Status (Current/Former) (Phase 1)
- Final Gold Baseline
- BMI
- Diabetes (Yes or No)
- High Blood Pressure (Yes or No)
- High Cholesterol (Yes or No)
- Coronary Artery Disease (Yes or No)

compared the obesity measures across high-risk subtypes of emphysema and/or airway disease. The PCAsubtype developed (Kinney) high risk for mortality. mphysema is when the alveoli (air sacs) in the lungs are compromised (NCI) sthma is very similar to Airway Disease (Reactive). Airway Disease is brought on when a person's bronchial tubes (responsible for allowing air in the lungs) are swollen due to an irritant, and thus the person has breathing issues (Reactive). The PCAsubtypes

1. High-Risk Airway Disease High-Risk Emphysema, combination of High-Risk Airway Disease & High-Risk Emphysema and no subtypes

Researchers have found a way to predict two traits High Risk Emphysema High-Risk Airway Diseaseis done by PCAsubtype. We are looking to see if PCAsubtype can predict other COPD traits. e that there is a link between Airway Disease and obesity. this hypothesis 65% of people diagnosed with COPD are obese (Young). study COPD from the viewpoint of the subtypes and five measures. In this paper we will be specifically studying the PCAsubtype that includes the High-Risk Airway Disease group.

The adiposity measures of interest for this project are listed as follows:

1. Abdominal visceral fat (ABVF)
2. Subcutaneous fat (SFA)
3. Pectoralis Muscle Area (PMA)
4. Lean PMA (PMAlean)
5. Liver (liverdensity)

Statistical Analysis:

Summary Statistics for each individual variable generated frequency tables or box plots as appropriate. Next, one-way ANOVA tables modeling each of the 10 different combinations of adiposity measurements and the PCAsubtype. Originally, to include the ANOVA tables with the Tukey Post-Hoc. However, the normality was violated, so Kruskal-Wallis & the Wilcoxon Rank Sum Test. The appropriate Wilcoxon Rank Sum Test with continuity correction to determine which groups different. All statistics were computed using the statistical computing software R version 4.2.0 using a significance level of 0.05.

The following output data were produced for the client:

- A frequency table for the final gold variable, newHRsubtype, and the PCAsubtype
- A percentage table for the final gold, newHRsubtype, and the PCAsubtype.
- Tables for each type of adiposity measurement categorized via the newHRsubtype and the PCAsubtype to create an average table.
- A statistical analysis of the obesity measures across moderate and high-risk subtypes of emphysema and/or airway disease
- Other needed statistical results as requested by the client

· A compilation of all the work into a format to be inserted into a grant proposal being prepared by Dr. Young.

Analysis & Results

Table 1. Distribution of Study Population by Diabetes Table via (Frequency & Percentages):

We can see that we have 705 more males than females in this study. There are more Males than Females with diabetes. However, there are overall more male participants than Female participants.

	(Percentage of Males:	of Females:	Total:
:	782 (7.67%)	555 (5.44%)	1,337
:	4,670 (45.79%)	4,192 (41.10%)	8,862

Table 2. Mean BMI ± Standard Deviation (Confidence Intervals) at Phase 1 by :

The average BMI and its standard deviation among Females are slightly higher than for Males.

:	Mean BMI ± Standard Deviation at Phase 1 (Confidence Intervals):
Male:	28.45 ± 5.62
Female:	29.26724 ± 6.937213

Table 3. & High Blood Pressure (HBP) Table:

More males reported high blood pressure than did females.

	No HBP Frequency Percentage):	HBP Reported
Male:	3,138 (30.77%)	2,313 (22.68%)

Female:	2,662 (26.10%)	2,085 (10.45%)

The Final Gold Baseline Variable is a measure of the disease severity of COPD. COPD was grouped as spirometric grades 1-4 based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines (Rabe KF). Participants without spirometric evidence of airflow obstruction ($FEV_1/FVC \geq 0.70$ and $FEV_1 \geq 80\%$ predicted) were classified as GOLD 0. Subjects with $FEV_1/FVC \geq 0.70$ and $FEV_1 <80\%$ predicted were classified as Preserved Ratio Impaired Spirometry (PRISm) (Wan ES). Here we can see a frequency table for the Final Gold Baseline variable. We can see that most people belong to the control section of this variable.

- Control: (FEV1 >= 80%, FEV1/FVC >= 0.7)
- GOLD 1: (FEV1 >= 80%, FEV1/FVC < 0.7)
- GOLD 2: (50% <= FEV1 < 80%, FEV1/FVC < 0.7)
- GOLD 3: (30% < = FEV1 < 50%, FEV1/FVC < 0.7)
- GOLD 4: (FEV1 < 30%, FEV1/FVC < 0.7)

- Prism: (Preserved Ratio Impaired Spirometry) (FEV1/FVC >= 0.7 but FEV1 < 80%)

Here we can see a percentage table for Final Gold Baseline variable. We can see that the highest percentage belongs to the control section of this variable.

Table 4. Final Gold Baseline Variable Frequency & Percentage:

	Prism:	Control:	GOLD 1:	GOLD 2:	GOLD 3:	GOLD 4:	NA:
Frequency:	1,262	4,388	787	1,926	1,164	606	66
Percentage:	12.374%	43.024%	7.716%	18.884%	11.412%	5.942%	0.647%

We can see that the highest number of people, by count and percent, appear in the Normal category of this variable. There are also a lot of NA values associated with this variable. NA values by including them in the table so that we could see how many NA values we are working with here. We can also see that a bit over 20% of all values of this variable are not available.

Table 5. Frequency & Percentage of PCAsubtype:

	Normal:	HR Airway w/o Emph:	HR Airway w/ Emph:	Emph w/o HR Airway:	NA:
Frequency:	5,519	1,007	625	1,006	2,042
Percentage:	54.11%	9.87%	6.13%	9.86%	20.02%

Result Section 1. Abdomen Visceral Fat (ABVF):

Result Section Abdomen Visceral Fat.

We know that the mean is the average of a data set. The means are rather high here, meaning the expectation will be higher. The High-Risk Airway Disease without Emphysema group had the highest mean of 179.0074 cm^2. When looking at Abdomen Visceral Fat (ABVF) categorized via the PCAsubtype variable Mean notice that the means of all of the groups of the PCAsubtype variable differ. he Emphysema without High-Risk Airway disease group has the smallest mean out of all the groups of the PCAsubtype variable having a value of 126.9193 cm^2.

bdomen visceral fat (ABVF) divided into subtypes via the PCAsubtypes variable standard deviation is shown below. We know that the standard deviation is a measure of how dispersed the data is in relation to the mean. Overall, each subtype has a rather high standard deviation, meaning that the data is more spread out. Once again, the High-Risk Airway Disease

Emphysema has the highest standard deviation out of all the subtypes with a standard deviation of 108.702 cm^2.

PCAsubtype:	Mean (AVBF)	Standard Deviation (AVBF)
Normal	147.2671	99.51597
High Risk Airway Disease w/o Emphysema	179.0074	108.70218
High Risk Airway Disease with Emphysema	141.5523	101.65187
Emphysema w/o HR Airway Disease	126.9193	87.22716

Box Plot:

Figure 1. Boxplot of the means of the groups of the PCAsubtype variable.

Note that the red dot represents the mean of each group of the PCAsubtype variable, and the boxes represent the middle 50%. When looking at the box plot, we want to see if there are differences in the means. When observing the box plot, we can see that there are definitely differences in the means among the categories of the PCAsubtype variable, but while there are differences there are not extreme differences in the means of the categories of the variable of interest. It appears that group 1 has the highest mean, and this might be driving the significant p-value for the Kruskal-Wallis Rank Sum Test. Notice that groups 2 = High Risk Airway Disease with Emphysema, 3 = Emphysema w/o HR Airway Disease, and the NA values all have somewhat similar means. pecifically, groups 3 and the NA seem to very simila. The -9 = Normal group has a higher spread compared to the other groups of the PCAvariable due to those outliers at the top.

Post Hoc Analysis tests were performed to determine which groups' mean differ from one another for this variable. The non-parametric Tukey Test was used to perform the Post-Hoc tests. Regarding the Post Hoc Analysis, if the P-Value is below 0.05 that implies that the difference in the mean of the certain groups is significantly different. From the results of the Post Hoc Analysis Tukey-Test (Simultaneous Test for General Linear Hypothesis) we can see from the p-value the following group's differences are significantly different:

- High Risk Airway Disease w/o Emphysema - Normal = 0

- Emphysema w/o HR Airway Disease - Normal = 0
- HR Airway with Emphysema – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0

Linear Hypothesis:	Estimate:	Standard Error:	T Value:	P-Value:
High Risk Airway Disease w/o Emphysema - Normal = 0:	31.740	3.531	8.990	**<0.001**
High Risk Airway Disease with Emphysema - Normal = 0:	-5.715	4.309	-1.326	0.5329
Emphysema w/o HR Airway Disease - Normal = 0:	-20.348	3.507	-5.802	**<0.001**

HR Airway with Emphysema – HR Airway w/o Emphysema = 0:	-37.455	5.214	-7.184	**<0.001**
Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0:	-52.088	4.573	-11.391	**<0.001**
Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0:	-14.633	5.198	-2.815	**0.0238**

The Kruskal-Wallis Test is a nonparametric tool that can determine if distributions are different. A nonparametric alternative to a one-way ANOVA table is the Kruskal-Wallis Test (Kruskal-Wallis). We used the Kruskal-Wallis Test as we obtained violations of the ANOVA. Note that the Kruskal-Wallis Test cannot tell us exactly which groups are statistically significant, it will only tell us that at least two groups are different (Kruskal-Wallis). Since the P-Value is smaller than the alpha value of 0.05 we know that at least two groups are different, i.e., statistically significant.

Kruskal-Wallis Chi-Squared:	131.27
Degrees of Freedom:	3
P-value:	< 2.2e-16

The Wilcoxon Rank Sum Test is a nonparametric test serving as an alternative test to the independent sample t test (Ellis). To determine if the mean of the sum of ranks (medians) of two groups are statistically different, the Wilcoxon Rank Sum Test can be used (Ellis). From the results of the Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction we can see that all groups that were compared except the High-Risk Airway with Emphysema compared to the Normal group differ significantly from each other as all of the p-values are below the alpha level of 0.05. Note that the High Risk Airway with Emphysema compared to the Normal group does not differ significantly as the p-value obtained here is greater than the alpha level of 0.05.

:

	Normal:	HR Airway w/o Emphysema:	HR Airway w/ Emphysema:
HR Airway w/o Emphysema:	< 2e-16	NA	NA
HR Airway w/ Emphysema:	0.062	2.3e-12	NA
Emphysema without	1.8e-08	< 2e-16	0.043

HR Airway Disease:			

Result Section 2. Subcutaneous Fat Area (SFA):

In the Result Section we will examine the Subcutaneous Fat Area (SFA).

When observing the Standard Mean Table of Subcutaneous Fat Area (SFA) categorized via the PCAsubtypes variable we can see that numerically the values of the means of each group do differ. Notice that the mean of the Normal group of the PCAsubtype variable is the highest among all of the groups of PCAsubtype variable of 67.89439. Also note that the smallest mean of all of the groups of the PCAsubtype variable High Risk Airway Disease with Emphysema value f 45.75640 cm^2.

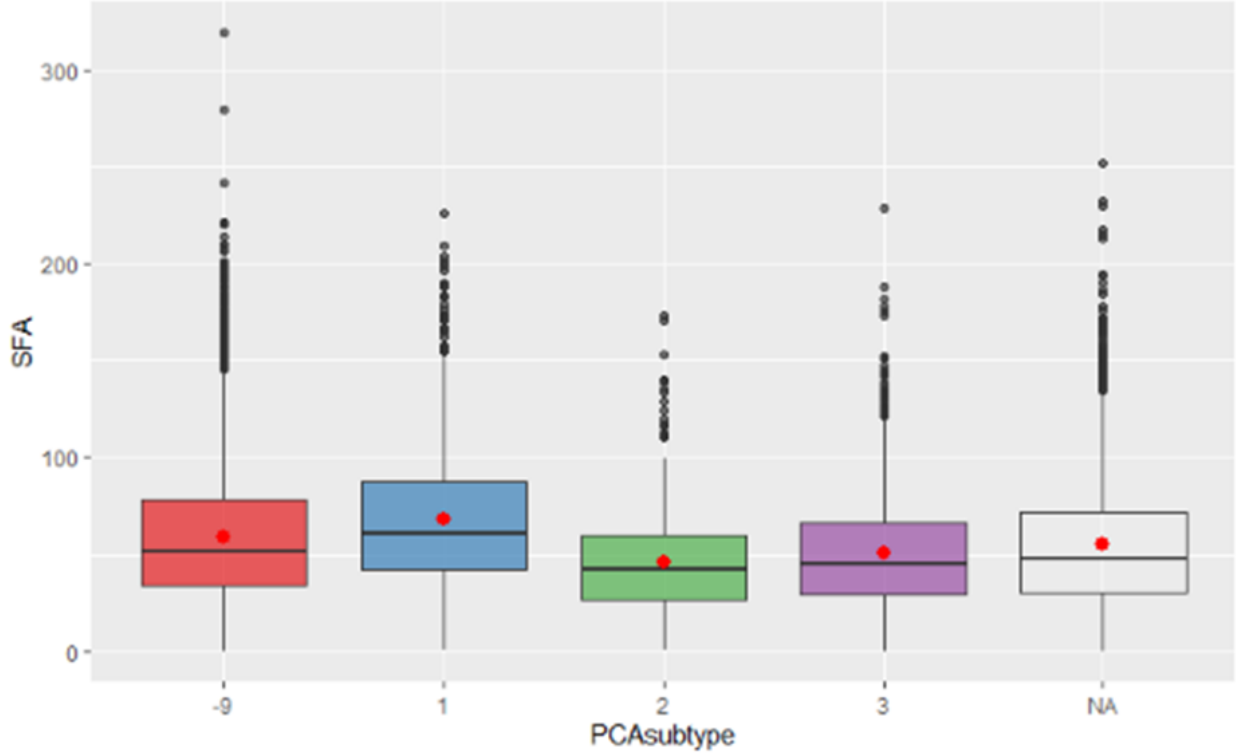
he standard deviations among the groups of the PCAsubtype variable o differ. Notice that the High Risk Airway Disease without Emphysema group of the PCAsubtype variable ha the highest standard deviation 37.25602 cm^2. Also note that the High Risk Airway Disease with Emphysema group has the smallest standard deviation in the SFA setting. Likewise, the High Risk Airway Disease without Emphysema group also had the smallest mean in the SFA setting; the High Risk Airway Disease without Emphysema group also had the smallest mean in the AVBF setting.

PCAsubtype	Mean (SFA):	Standard Deviation (SFA):

Normal	59.17342	36.01395
High Risk Airway Disease w/o Emphysema	67.89439	37.25602
High Risk Airway Disease with Emphysema	45.75640	26.62793
Emphysema w/o HR Airway Disease	50.79913	30.63772

Box Plot:

Recall that the red dot shown on the box plot represents the mean of each group of the PCAsubtype variable. The first thing to note about the box plot is that the mean of the group 1 = High Risk Airway Disease without Emphysema corresponds to the largest mean among all of the groups. The high mean value of group 1 could potentially be the driving factor for the significant P-value obtained for the Kruskal-Wallis Rank Sum Test. Our main goal when observing the box plot is to see if there are any differences in the means of each group of the PCAsubtype variable. When studying the box plot above we can see that there are definitely differences in the means of each group, however the differences among the means of the groups are not extremely different. Note that the means of the 2 = High Risk Airway Disease with Emphysema, 3 = Emphysema, and NA groups seem to have somewhat similar means, especially group 3 and the NA group appearing to have very similar means. Notice that the – 9 = Normal groups obtains the highest spread compared to the other groups as this groups contains a lot of outliers.



-9 = Normal
1 = High Risk Airway Disease w/o Emphysema
2 = High Risk Airway Disease with Emphysema
3 = Emphysema w/o HR Airway Disease

We performed Post Hoc Analysis here as we needed to determine which groups means differ among one another. Once again, the non-parametric Tukey Test was used to perform the post-hoc analysis. Recall that if the p-value obtained in the post hoc analysis test is below the chosen alpha value of 0.05, the difference in the means of the two groups is statistically significant. From the results obtained from the Post Hoc Analysis Tukey-Test (Simultaneous Test for General Linear Hypothesis) we can see from the p-value being less than the alpha value of 0.05 that the following group's differences are statistically significant:

- High Risk Airway Disease w/o Emphysema - Normal = 0
- High Risk Airway Disease with Emphysema - Normal = 0
- Emphysema w/o HR Airway Disease - Normal = 0
- HR Airway with Emphysema – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0

Thus, all of the differences observed were statistically significant, all of the means differing.

Table 11. Post Hoc Analysis on the One Way ANOVA of the SFA as a Function of the PCAsubtype(Simultaneous Test for General Linear Hypotheses):

Linear Hypothesis:	Estimate:	Standard Error:	T Value:	P-Value:

High Risk Airway Disease w/o Emphysema - Normal = 0:	8.721	1.221	7.141	**<0.001**
High Risk Airway Disease with Emphysema - Normal = 0:	-13.417	1.502	-8.931	**<0.001**
Emphysema w/o HR Airway Disease - Normal = 0:	-8.374	1.216	-6.884	**<0.001**
HR Airway with Emphysema – HR Airway w/o Emphysema = 0:	-22.138	1.814	-12.202	**<0.001**
Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0:	-17.095	1.586	-10.780	**<0.001**
Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0:	5.043	1.811	2.784	**0.0257**

Notice that the degrees of freedom are 3 as we would expect as there four different groups in the PCAsubtype variable. When observing the P-value obtained from the Kruskal-Wallis Rank Sum Test which is < 2.2e-16, we know that at least two groups of the PCAsubtype variable obtain different mean values, i.e, statistically significant.

Kruskal-Wallis Chi-Squared:	197.82
Degrees of Freedom:	3
P-value:	**< 2.2e-16**

When observing the Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction, we can see that all of the groups of the PCAsubtype variable that were compared differed significantly from each other as we can see that every p-value obtained is below the alpha level of 0.05. This is no surprise all of the groups have been shown to be statistically significant in all of the tests thus far.

	Normal	HR Airway w/o Emphysema	HR Airway w/ Emphysema
HR Airway w/o Emphysema	**5.3e-14**	NA	NA
HR Airway w/ Emphysema	**< 2e-16**	**2.3e-16**	NA
Emphysema without HR Airway Disease	**1.2e-10**	**< 2e-16**	**0.0053**

Result Section 3. Pectoralis Muscle Area (PMA):

In Result section we will examine the Pectoralis Muscle Area (PMA). Pectoralis Muscle Area (PMA) measures muscle area, with lower values meaning lower muscle mass.

Observing the Pectoralis Muscle Area (PMA) categorized by the PCAsubtypes Standard Mean we can instantly see that the means of the four different groups of the PCAsubtype variable differ numerically. The largest mean among the four different groups of the PCAsubtype variable is the Normal group with a mean of 44.30675 cm^2. Also, note that the High-Risk Airway Disease with Emphysema groups corresponds to the smallest mean with a mean of 31.86294 cm^2.

Recall that standard deviation is highly related to the mean the calculation of standard deviation is dependent on the mean. When observing the Pectoralis Muscle Area (PMA) categorized via the PCAsubtypes variable Standard Deviation table we can see that the standard deviations of the different groups of the PCAsubtype differ, meaning that the means of the groups differ as we have previously seen. Note that the highest standard deviation corresponds to the Normal group a standard deviation of 17.14181 cm^2. Recall that the Normal group also had the highest mean. Also, note that the smallest standard deviation corresponds to the High-Risk Airway Disease with Emphysema group at a value of 11.65232 cm^2. Likewise, the High-Risk Airway Disease with Emphysema group also had the smallest mean in the PMA setting among the groups as previously found in the SFA setting.

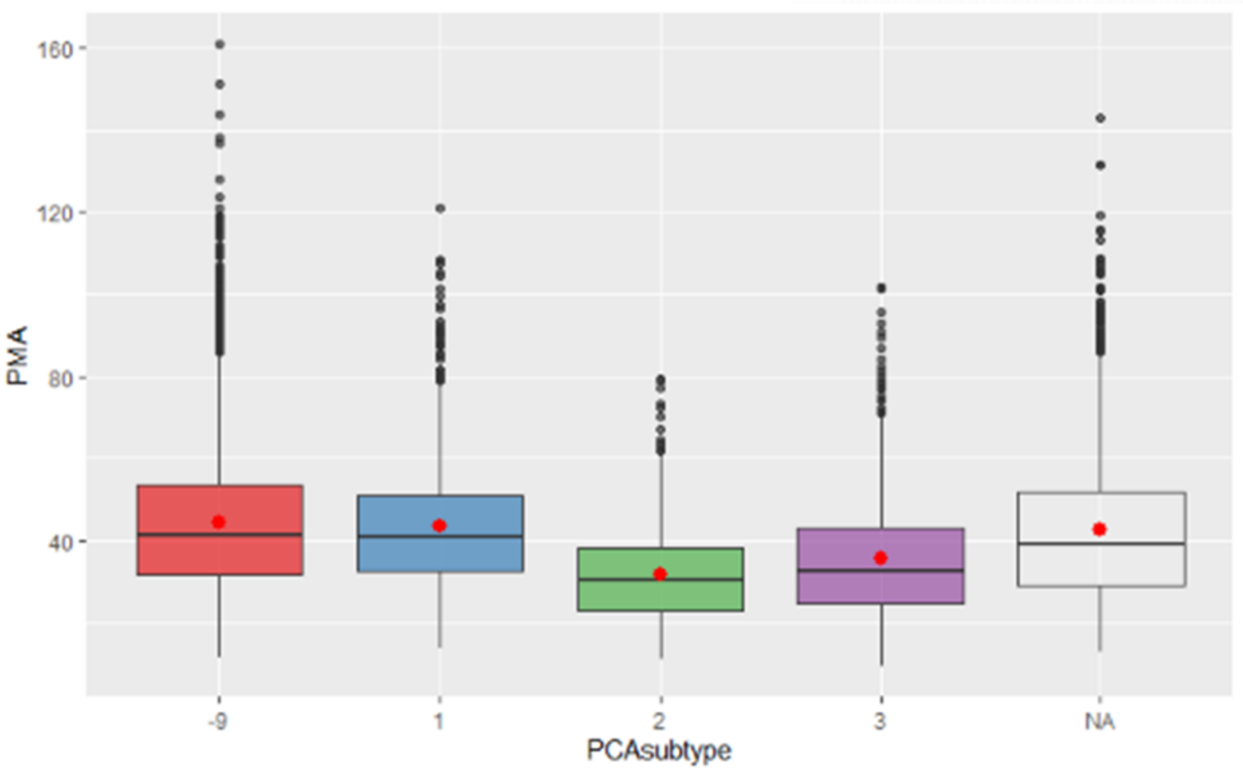
	Mean (PMA)	Standard Deviation (PMA)
Normal	44.30675	17.14181

High Risk Airway Disease w/o Emphysema	43.32750	15.58320
High-Risk Airway Disease with Emphysema	31.86294	11.65232
Emphysema w/o HR Airway Disease	35.50040	14.71625

Box Plot:

Boxplot of the means of the groups of the PCAsubtype variable.

Note that the red dot shown on the box plot represents the mean of each group of the PCAsubtype variable. First, we can see that the means of the Normal, High-Risk Airway Disease without Emphysema, and the NA groups of the PCAsubtype variable seem to be very similar. somewhat hard to determine which groups of the PCAsubtype variable has the largest means among the groups that the Normal group the largest mean . This high mean value of group could be the driving factor for obtaining the very small p-value when conducting the Kruskal-Wallis Rank Sum Test. Note that our main goal when observing the box plot is to see if there any visual differences among the means of the groups of the PCAsubtype variable. When studying the box plot, we can see that there are definitely differences in the mean the groups. We can see that the means of the Normal, High Risk Airway Disease without Emphysema, and the NA groups differ from the means of the High Risk Airway Disease with Emphysema and Emphysema without High-Risk Airway Disease. Finally, notice that the Normal groups appears to obtain the highest spread compared to the other PCAsubype as this group contains a lot of outliers.



-9 = Normal
1 = High Risk Airway Disease w/o Emphysema
2 = High Risk Airway Disease with Emphysema
3 = Emphysema w/o HR Airway Disease

Post Hoc Analysis was performed to determine which groups means actually differ from one another. We used the non-parametric Tukey Test perform the Post-Hoc analysis. Recall that when the p-value obtained in the post hoc analysis test is below the chosen alpha level of 0.05, the difference in the means of the difference of the two groups is statistically significant. From the results obtained from the Post Hoc Analysis Tukey-Test (Simultaneous Test for General Linear Hypothesis) we can see that the following group's differences are statistically significant as the p-value is less than the alpha value of 0.05:

- High Risk Airway Disease with Emphysema - Normal = 0
- Emphysema w/o HR Airway Disease - Normal = 0
- HR Airway with Emphysema – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0

· Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0

Thus, all of the differences observed the High Risk Airway Disease w/o Emphysema -Normal = 0 are statistically significant, meaning all the means differ.

Linear Hypothesis	Estimate	Standard Error	T Value	P-Value
High Risk Airway Disease w/o Emphysema - Normal = 0	-0.9792	0.5669	-1.727	0.298
High Risk Airway Disease with Emphysema -Normal = 0	-12.4438	0.6977	-17.834	**<1e-04**
Emphysema w/o HR Airway Disease - Normal = 0	-8.8063	0.5661	-15.555	**<1e-04**
HR Airway with Emphysema – HR Airway w/o Emphysema = 0	-11.4646	0.8423	-13.612	**<1e-04**
Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0	-7.8271	0.7369	-10.621	**<1e-04**

Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0	3.6375	0.8418	4.321	**<1e-04**

First, notice that the degrees of freedom here is 3 as there are four groups of the PCAsubtype variable. The Kruskal-Wallis Test is a nonparametric tool that can determine if distributions are different. Recall that the Kruskal-Wallis Rank Sum Test is used to see if any of the means of the groups of the PCAsubtype variable are different, i.e., statistically significant. When observing the p-value of < 2.2e-16 obtained from the Kruskal-Wallis Rank Sum Test, this implies that at least two groups of the PCAsubtype variable are different, i.e., statistically significant.

Kruskal-Wallis Chi-Squared	570.81
Degrees of Freedom	3
P-Value	**< 2.2e-16**

When observing the results from the Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction, we see that the means of all of the groups of the PCAsubtype variable except the High-Risk Airway without Emphysema group compared to the Normal that differ significantly from each other as the obtained p-values are all very small, below the chosen alpha level of 0.05. This supports all of our previous results as the same groups have appeared significantly significant in all of the other tests that have been completed.

	Normal	HR Airway w/o Emphysema	HR Airway w/ Emphysema
HR Airway w/o Emphysema	0.29	NA	NA
HR Airway w/ Emphysema	**< 2e-16**	**< 2e-16**	NA
Emphysema without HR Airway Disease	**1.2e-10**	**< 2e-16**	**2.3e-05**

Result Section 4. Lean Pectoralis Muscle Area (PMAlean):

In Result section 4 we will examine the Pectoralis Muscle Area (PMAlean).

When observing the Lean Pectoralis Muscle Area (PMAlean) categorized by the groups of the PCAsubtype 8, we can see that the means of the four different groups differ numerically, none of the groups have the same mean value. We can see that the largest mean value of 41.48651. the smallest mean value 30.63551 cm^2.

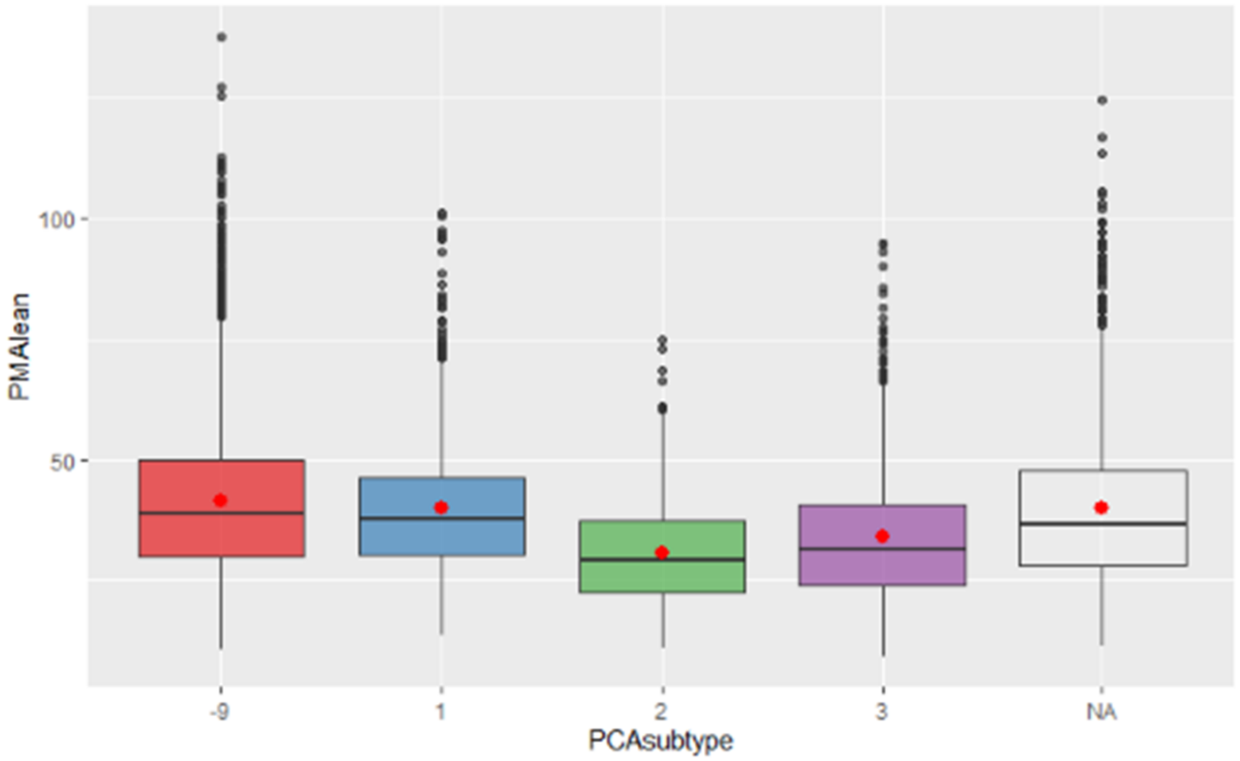
We know that the standard deviation is highly dependent on the mean. When observing the Lean Pectoralis Muscle Area (PMAlean) categorized by the groups of the PCAsubtype Standard Deviation Table we can see that the standard deviations of the different groups of the

PCAsubtype variable differ. This corresponds to what we saw in the mean table with each group of the PCAsubtype variable having a different mean. Note that the largest standard deviation belongs to the Normal group with a value of 15.51217. Also note that the smallest standard deviation value corresponds to the High-Risk Airway Disease with Emphysema group at a value of 10.72845 cm^2.

PCAsubtype:	Mean (PMAlean):	Standard Deviation (PMAlean)
Normal:	41.48651	15.51217
High Risk Airway Disease w/o Emphysema:	39.90604	13.67054
High Risk Airway Disease with Emphysema:	30.63551	10.72854
Emphysema w/o HR Airway Disease:	33.89474	13.53317

Box Plot:

he red dot shown on the box plot represents the mean of each group. When observing the plotted box plot, we see that the Normal, High Risk Airway Disease without Emphysema, and the NA groups all appear to have somewhat similar means. This corresponds to what we found in our Mean table of the Normal group having the largest mean among the groups. The high mean n Normal group could be the reason that we obtained the significant p-value from the Kruskal-Wallis Rank Sum Test. Recall that the main goal when observing the box plot is to see if there are any visual differeces in the means of the groups of the PCAsubtype variable. When studying the box plot we can see that the means of the groups definitely do differ. Notice that the High Risk Airway Disease with Emphysema group appears to have the smallest mean value among the group, which to our finding earlier with group 2 having the smallest mean value numerically. Finally, notice that the Normal group the highest spread compared to the other groups group contains a lot of outliers.



-9 = Normal
1 = High Risk Airway Disease w/o Emphysema
2 = High Risk Airway Disease with Emphysema
3 = Emphysema w/o HR Airway Disease

Post Hoc analysis was performed as we needed to determine exactly which groups' means are statistically significant. As previously performed, we used the non-parametric Tukey Test while performing the post hoc analysis. Recall that if the p-value obtained from the post hoc analysis is below the chosen alpha level of 0.05, the difference between the two group’s means is statistically significant. When observing the results of the Post Hoc Analysis Tukey-Test (Simultaneous Test for General Linear Hypothesis) we can see that the difference of the following groups’ p-value was smaller than the chosen level of alpha (0.05):

- High-Risk Airway Disease w/o Emphysema - Normal = 0
- High-Risk Airway Disease with Emphysema - Normal = 0
- Emphysema w/o HR Airway Disease - Normal = 0
- HR Airway with Emphysema – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0

Thus, all of the differences between groups observed here were statistically significant, i.e., all of the mean values are different.

Linear Hypothesis:	Estimate:	Standard Error:	T Value:	P-Value:
High Risk Airway Disease w/o Emphysema - Normal = 0:	-1.5805	0.5124	-3.084	**0.0102**
High Risk Airway Disease with Emphysema - Normal = 0:	-10.8510	0.6307	-17.203	**<0.001**
Emphysema w/o HR Airway Disease - Normal = 0:	-7.5918	0.5118	-14.834	**<0.001**
HR Airway with Emphysema – HR Airway w/o Emphysema = 0:	-9.2705	0.7614	-12.176	**<0.001**
Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0:	-6.0113	0.6662	-9.024	**<0.001**
Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0:	3.2592	0.7609	4.283	**<0.001**

When observing the results of the Kruskal-Wallis Rank Sum Test we can first note that the degrees of freedom is 3, which is what we would expect as there are 4 groups of the

PCAsubtype variable. Note that the p-value obtained from the Kruskal-Wallis is very small at a value of < 2.2e-16, which is smaller than the chosen alpha level of 0.05. Since the obtained p-value is smaller than the chosen alpha level of 0.05, we know that at least two groups of the PCAsubtype different mean values, i.e., statistically significant differences between two or more of the groups appear.

Kruskal-Wallis Chi-Squared	501.65
Degrees of Freedom	3
P-value	**< 2.2e-16**

When observing the Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction, we see that all of the groups of the PCAsubtype variable that are compared to one another differed significantly from one another as every p-value obtained here is smaller than the chosen alpha level of 0.05. This is what we would expect as every group of the PCAsubtype variable appeared statistically significant in all of the other tests that were completed.

	Normal	HR Airway w/o Emphysema	HR Airway w/ Emphysema
HR Airway w/o	**0.023**	NA	NA

Emphysema			
HR Airway w/ Emphysema	**< 2e-16**	**< 2e-16**	NA
Emphysema without HR Airway Disease	**< 2e-16**	**< 2e-16**	**5.4e-05**

Result Section 5. Liver Density:

In Result section we will examine the Liver Density.

Recall that our null hypothesis is that the means of each of the groups are the same. We can see from the Liver Density categorized by the groups of the PCAsubtype Standard Mean Table that the mean of each group is different. Notice that the largest mean corresponds to the Emphysema without High Risk Airway Disease group with a value of 58.06102 Hounsfeld (Hu). Also notice the smallest mean corresponds to the High Risk Airway Disease with Emphysema group with a value of 54.11338.

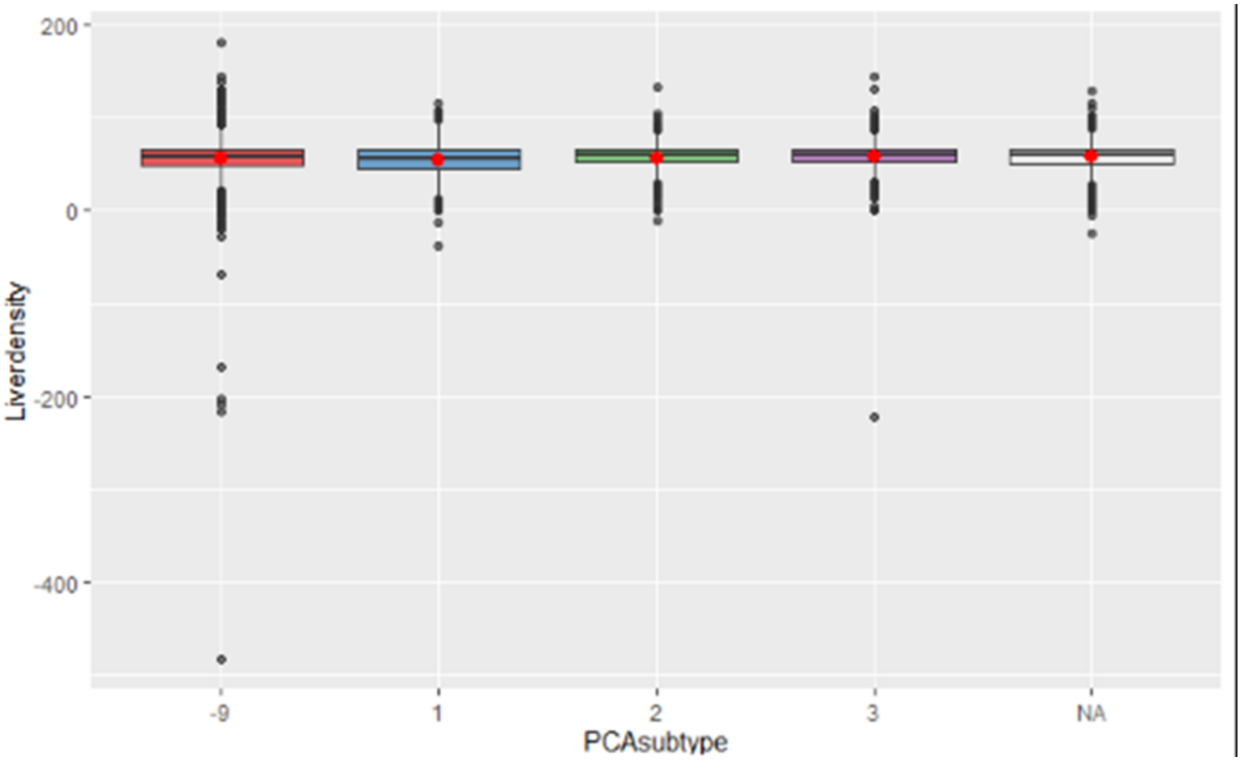
We know that the standard deviation is heavily related to the mean. When studying the Liver Density categorized by the groups of the PCAsubtype Table we can see that the standard deviations of the PCAsubtype are all different. Note that the largest standard deviation corresponds to the Normal group with a value of 20.44485 Hu. This is as we found the group with the largest mean value to be the Emphysema without High-Risk Airway Disease. Also note that the High-Risk Airway Disease with Emphysema corresponds to the smallest standard deviation with a value of 15.63679. This is unusual we previously found that Emphysema without High-Risk Airway group has the largest mean y, we would think that it would also have the largest standard deviation. Therefore, we know that the means of the groups of the PCAsubtype are not all the samethere is a statistically significant difference between the mean of the four PCAsubtype.

PCAsubtype	Mean	Standard Deviation
Normal	55.87894	20.44485
High Risk Airway Disease w/o Emphysema	54.11338	17.99885
High Risk Airway Disease with Emphysema	57.16758	15.63679
Emphysema w/o HR Airway Disease	58.06102	17.75758

Box Plot

Boxplot of the means   the PCAsubtype .

he red dot shown on the box plot corresponds to the mean value. Recall that the main thing we are looking for when observing the box plot is any visual differences in the means   the PCAsubtype s. e can see that there are differences in the mean,differences . The means of the groups actually look pretty similar, and it is somewhat difficult to determine which group the largest mean value. Notice that theormal group the spread compared to the other PCAsubtype as the Normal group contains a lot of outliers. We can see that overall, the meanof the PCAsubtype are fairly large as the red dot is very high on the graph.



-9 = Normal

1 = High Risk Airway Disease w/o Emphysema
2 = High Risk Airway Disease with Emphysema
3 = Emphysema w/o HR Airway Disease

Post Hoc analysis was performed as we needed to determine exactly which means of which group were different. The Tukey Test was used while performing the post hoc analysis. Recall that if the p-value obtained from the post hoc analysis is smaller than the chosen alpha value, then the difference means of the two groups is statistically significant. Observing the results obtained from the Post Hoc Analysis (Simultaneous Test for General Linear Hypotheses) we can see that the following groups are associated with a p-value smaller than the chosen alpha level of 0.05:

- High Risk Airway Disease w/o Emphysema - Normal = 0
- Emphysema w/o HR Airway Disease - Normal = 0
- HR Airway with Emphysema – HR Airway w/o Emphysema = 0
- Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0

Thus, all of the difference means were statistically significant, except the following :

- High Risk Airway Disease with Emphysema - Normal = 0

- Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0

Linear Hypothesis	Estimate	Standard Error	T Value	P-Value
High Risk Airway Disease w/o Emphysema - Normal = 0	-1.76557	0.6691	-2.639	**0.03881**
High Risk Airway Disease with Emphysema - Normal = 0	1.2886	0.8239	1.5464	0.38621
Emphysema w/o HR Airway Disease - Normal = 0	2.1821	0.6693	3.260	**0.00586**
HR Airway with Emphysema – HR Airway w/o Emphysema = 0	3.0542	0.9941	3.072	**0.01050**
Emphysema w/o HR Airway – HR Airway w/o Emphysema = 0	3.9476	0.8703	4.536	**<0.001**
Emphysema w/o HR Airway – HR Airway w/ Emphysema = 0	0.8934	0.9943	0.899	0.79703

When observing the results of the Kruskal-Wallis Rank Sum Test we can see that the degrees of freedom is 3, which is what we would expect as the PCAsubtype variable has four different groups. Notice that the p-value obtained from the Kruskal-Wallis Rank Sum Test is 1.321e-08, which smaller than the chosen alpha level of 0.05. Thus, since the p-value obtained from the Kruskal-Wallis Rank Sum Test is smaller than the chosen alpha level, we know that at least two groups of the PCAsubtype variable different mean values.

Kruskal-Wallis Chi-Squared	39.56
Degrees of Freedom	3
P-value	**1.321e-08**

When observing the Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction,of the groups compared a very small p-value, smaller than the chosen alpha level of 0.05. This corresponds to our previous results from the Post-Hoc analysis the difference between the High-Risk Airway Disease with Emphysema group and the Normal groups & groups Emphysema without High-Risk Airway Disease and High Risk Airway Disease with Emphysema did not show any significance.

	Normal	HR Airway w/o Emphysema	HR Airway w/ Emphysema
HR Airway w/o Emphysema	**0.023**	NA	NA
HR Airway w/ Emphysema	**< 2e-16**	**< 2e-16**	NA
Emphysema without HR Airway Disease	**< 2e-16**	**< 2e-16**	**5.4e-05**

Shown below able all of the p-values obtained when using the Kruskal-Wallis test on each of the five different of interest. We can see that every single p-value obtained is significant as the p-value for each is smaller than the chosen alpha value of 0.05.

	P-Value (Kruskal-Wallis)
Visvercal Fat (ABVF)	**< 2.2e-16**
Subcutaneous Fat (SFA)	**< 2.2e-16**
Pectoralis Muscle Area (PMA)	**< 2.2e-16**

Lean PMA (PMAlean)	**< 2.2e-16**
Liver (liverdensity)	**1.321e-08**

Discussion:

Overall, we observed some very significant results. In each of the five different scenarios (the five different types of fat categorized by the 4 different groups of PCAsubtype variable) we obtained statistically significant results. Thus, in each of the five scenarios we know that the means of the groups of the PCAsubtype are not the same, they differ significantly. In the first scenario where we were studying the Abdomen Visceral Fat (ABVF) categorized by the PCAsubtype we obtained a significant p-value when performing the Kruskal-Wallis Rank Sum Test and that all the difference means were significant the difference of the High-Risk Airway Disease with Emphysema - Normal = 0 difference. In the second scenario Subcutaneous Fat Area (SFA) categorized by the four PCAsubtype we obtained a significant p-value when performing the Kruskal-Wallis Rank Sum Test Post Hoc analysis that all the groups had different mean. Now, in the third scenario the Pectoralis Muscle Area (PMA) categorized by the PCAsubtype we once again obtained a significant p-value when performing the Kruskal-Wallis Rank Sum Test, and we found that every difference of the two groups means e statistically significant except the High-Risk Airway Disease without Emphysema - Normal = 0 groups. Furthermore, in the fourth scenario (the lean pectoralis Muscle area categorized by the four groups of the PCAsubtype variable) we again obtain significant p-values when performing the Kruskal-Wallis Rank sum test and observed that all of the differences between the two groups were statistically significant, meaning that they had different mean values. Finally, in the fifth scenario (the liver density categorized by the four groups of the PCAsubtype) we acquired a significant p-values when performing the Kruskal-Wallis Rank Sum Test, and found from the post hoc analysis that difference groups observed was found to e significant difference, with the exception of the High Risk Airway Disease with Emphysema - Normal = 0 and the Emphysema without High Risk Airway Disease – High Risk Airways with Emphysema = 0. Thus, in nearly every single scenario performed we procured significant p-values m the Kruskal-Wallis Rank Sum Test and found significant differences in mean values among the different groups of the PCAsubtype variable.

My client’s goals were met as I completed the following s:

· A frequency table for the final variable & PCAsubtype

· Average & standard deviation tables for each type of adiposity measurement categorized via the PCAsubtype

· A statistical analysis of the obesity measures across moderate and high-risk subtypes of emphysema and/or airway disease

· A compilation of all the work was put into a format to be inserted into a grant proposal being prepared by Dr. Young.

Next steps for this analysis is to work the grant proposal.

Conclusions:

For this consulting project, we inspected the mean anthropometric measures among the subtypes of interest.

Model & Methods:

The first thing I did was import the quantitative data and clean it up. The data consists of 73 variables. One of the first things I did in R was select the following variables of interest that way we were working only with the variables of interest:

- Age Phase 1
- Gender
- Race
- Do the participants smoke cigarettes (as of one month ago) (Phase 1)
- Final gold Baseline
- BMI
- Diabetes (yes or no)
- High Blood Pressure (Yes or no)
- Coronary Artery Disease (yes or no)
- Pack per years

Recall that the main five fat measures of interest are as follows:

· Visceral Fat (ABVF)

· Subcutaneous Fat (SFA)

· Pectoralis Muscle Area (PMA)

· Lean PMA (PMAlean)

· Liver Density (Liverdensity)

Note that the main five fat measures of interest were categorized by the following subtype:

- High Risk Airway Disease, High Risk Emphysema, combination, and no subtypes

The methods used are listed as follows:

· Imported and cleaned up the data

· Standard Mean tables of the five different fats of interest categorized by the PCAsubtype

· Standard Deviation tables of the five different fats of interest categorized by the PCAsubtype

· One-way ANOVA Tables modeling the five different fats of interest as a function of the PCAsubtype

· Post Hoc Analysis (Simultaneous Tests for General Linear Hypothesis & Simultaneous Confidence Intervals) for each ANOVA table created

· Assessed the equality of variances by performed Levene’s Test for Homogeneity of Variance for each fat of interest with the PCAsubtype

· Tested whether samples originated from the same distribution by performing the Kruskal-Wallis Rank Sum Test for the five different fats of interest as a function of the PCAsubtype

· Pairwise Comparisons using Wilcoxon Rank Sum Test with Continuity Correction

· Calculated pairwise comparisons between group levels with corrections for multiple testing using the Wilcoxon Rank sum test for the five different fats of interest and the PCAsubtype

` o The adjustment methods used was the Bonferroni Correction

· Created 10 Box plot of the means with the five different fats of interest as the y values and the PCAsubtype for the x values

· Tested each level of the subtypes for normality using the Shapiro Test

· Performed Nemenyi’s non-parametric all-pairs comparison test for Kruskal-type ranked data via the Nemenyi’s All-Pairs Rank Comparison Test for the five fats of interest the PCAsubtype

issing data values were dealt with by including the NA values in our tables & box plots. When creating tables with the data I would set the parameter NA to be set to “always”. This would include how many NA values were present in the particular variable. On the Box plots that were created in this analysis we can see the NA values plotted. The NA values could have been either included or not included. I decided to include the NA values so that we could see how many values we were actually missing. When computing the standard deviation, I set the na.rm paramter to be TRUE so that the NA values were not included the analysis here.

Also note that quantitative data was used for the statistical analysis.

After importing and cleaning the data, I created Standard Mean Tables of the five different fats of interest categorized by the four groups of the subtype of interest using the following formula for the mean:

$$mean = \frac{sum\ of\ the\ terms}{number\ of\ terms}$$

After the standard mean tables were created, I created the Standard Deviation Tables of the five different fats of interest categorized by the two different subtypes using the following formula for the standard deviation:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$$

Where

· σ = the population standard deviation

· N = the size of the population

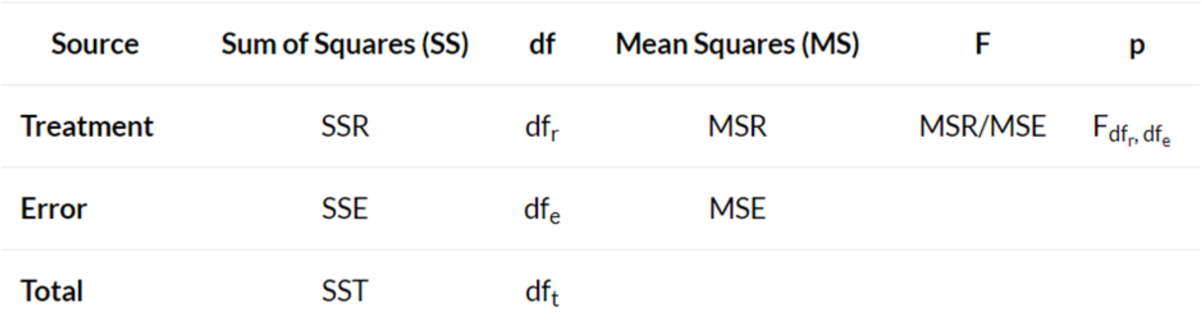
· x_i = each value from the population

· μ = the population mean

(Bhandari).

Note that N is dependent on the adiposity measures the PCAsubtype. Also, note that x_i is the certain adiposity measure. Furthermore, note that μ is dependent on the mean that was canulated.

After the Standard Deviation Tables were created, we created one-way ANOVA tables modeling the five different fats of interest as a function of the two different subtypes. The ANOVA table was created using the following formulas:



Source	Sum of Squares (SS)	df	Mean Squares (MS)	F	p
Treatment	SSR	df_r	MSR	MSR/MSE	F_{df_r, df_e}
Error	SSE	df_e	MSE		
Total	SST	df_t			

where:

· SSR: Regression Sum of Squares

· SSE: Error Sum of Squares

· SST: Total Sum of Squares (SST = SSR + SSE)

· df_r : Regression Degrees of Freedom ($df_r = k - 1$)

○ k : total observations

· df_t : Total Degrees of freedom ($df_t = n - 1$)

○ n : total observations

· df_e: Error Degrees of Freedom $df_e = n - k$

· MSR: Regression Mean Square ($MSR = \frac{SSR}{df_r}$)

· MSE: Error Mean Square ($MSE = \frac{SSE}{df_e}$)

· F: The F Test Statistic ($F = \frac{MSR}{MSE}$)

· P: The P-Value that corresponds to F_{df_r, df_e}

(Zach).

After the ANOVA tables were created, we ran the Post-Hoc analysis to determine exactly which groups of means differ from one another. The Post Hoc Analysis looks at the difference in the mean of each group. The estimate standard is the difference in the means of the two groups of interest. The Standard Error of the mean (SEM) is calculated as follows:

$$SEM = \frac{\sigma}{\sqrt{N}}$$

· SEM = The standard Error of the Mean

· σ = the population standard deviation

· N = the size of the population

(Tuovila).

Also note that the population standard deviation depends on the adiposity measurement the PCAsubtype.

The formula for the T-Value Test Statistic is as follows:

$$t = \frac{(\overline{x} - \mu_0)}{\frac{\sigma}{\sqrt{N}}}$$

· $\overline{x}$ = Sample Mean

· μ_0= Predicted Population Mean

· σ = the population standard deviation

· N = the size of the population

(Editor).

Now, the P-Value is obtained by utilizing the T-Distribution table noting that we will be using N-1 degrees of freedom (Zach).

found which differences of means were statistically different, we observed the Confidence Intervals of each difference in mean. To obtain the confidence intervals for each of the following differences in means we used the following formula:

$$CI = \bar{x} \pm \alpha \frac{\sigma}{\sqrt{N}}$$

· CI = Confidence Interval

· $\bar{x}$ = Sample Mean

· α = Confidence Level Value

· σ = the population standard deviation

· N = the size of the population

(Confidence Interval).

We performed the Kruskal-Wallis Rank Sum Test to test or not population samples originate from a different distribution. The Kruskal-Wallis Rank Sum Test was performed in the statistical consulting software R by using the kruskal.test function. The first statistical piece of information gained from the Kruskal-Wallis Rank Sum Test is the Kruskal-Wallis Chi-Squared value (also known as the H-Statistic) (Kruskal-Wallis) obtained by the following formula:

$$H = \left(\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right) - 3(n + 1)$$

· H = the Kruskal-Wallis Chi-Squared Value (H-Statistic)

· n = The sum of the sample sizes for all samples

· c = number of samples

· T_j = Sum of ranks in the jth sample

· n_j = size of the jth sample

(Kruskal-Wallis).

The degrees of freedom for the Kruskal-Wallis Rank Sum Test are obtained by subtracting one from c, $c - 1$ (Kruskal-Wallis). Then the critical chi-square value with an alpha level of 0.05 & $c - 1$ degrees of freedom (Kruskal-Wallis). he obtained Kruskal-Wallis Chi-Squared Value (H-statistic) is compared to the critical chi-square value. We know that the medians are not equal if the H-statistic is larger than the Kruskal-Wallis Chi-Squared value (Kruskal-Wallis). If the H-statistic is smaller than the Kruskal-Wallis Chi-squared value, then we conclude that the medians equal (Kruskal-Wallis).

We d from the Kruskal-Wallis test that there is a significant difference in the distributions of some groups. Thus, the Wilcoxon Rank Sum Test is performed to determine exactly which differences in the groups are significant. We calculated pairwise comparisons among group levels with corrections for multiple testing using the Wilcoxon Rank Sum Test. Recall that the method for adjusting pvalue was the Bonferroni Correction. The Pairwise comparisons using

Wilcoxon Rank Sum Test with continuity correction was performed in R using the pairwise.wilcox.test function, being sure to include the argument p.adjust.method = “BH”. The Bonferroni Correction is performed by dividing the chosen alpha value of 0.05 by the number of comparisons (Bonferroni). Thus, to obtain the Bonferroni Correction we use the following formula Note that the number of comparisons used here was 25 the chosen alpha value is 0.05. We reduced the chance of Type I error by using the Bonferroni Correction (Bonferroni). Thus,

- P(at least one significant result) = 1 – P(no significant results)

- P(at least one significant result) = 1 – $(1 - 0.05)^{25}$

- P(at least one significant result) = 0.72

After we determined which differences of means were significant, we created 10 box of the means of the   PCAsubtype . used the ggplot function in R to create the 10 different box plots. Note that the five adiposity measurements were used as the y values and the PCAsubtype for the x values. Each box plot shows 5 different boxes on the plot: one for each group of the PCAsubtype variable one box for the NA values. wanted to include the NA values so that we could see how the mean of the NA values compared to the means of the PCAsubtype. On each box a red dot plotted to represent the mean of each PCAsubtype. The plotted boxes represent the middle 50% of the data. The box plot also gives us other valuable pieces of information such as the minimum and maximum values of each PCAsubtype ; this also allows us to look for outliers. The box plots allow for us to easily visualize the data as well as visually see the means of the data, which is what we .

References:

- BMI (12/20/2021):

  - Centers for Disease Control and Prevention. (2021, August 27). *About adult BMI*. Centers for Disease Control and Prevention. Retrieved December 20, 2021, from https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html#:~:text=Body%20mass%20index%20(BMI)%20is,weight%2C%20overweight%2C%20and%20obesity.

- BMI (12/20/2021):

  - Centers for Disease Control and Prevention. (2021, September 30). *Adult obesity facts*. Centers for Disease Control and Prevention. Retrieved December 20, 2021, from https://www.cdc.gov/obesity/data/adult.html#:~:text=Obesity%20is%20a%20common%2C%20serious,from%204.7%25%20to%209.2%25.

- Confidence Interval (6/14/2022):

  - *Confidence interval: How to find it: The easy way!* Statistics How To. (2022, May 28). Retrieved June 14, 2022, from https://www.statisticshowto.com/probability-and-statistics/confidence-interval/.

- T-Value Statistic (6/14/2022):

- Editor, M. B. (n.d.). *One-sample t-test: Calculating the T-statistic is not really a bear*. Minitab Blog. Retrieved June 14, 2022, from https://blog.minitab.com/en/one-sample-t-test-calculating-the-t-statistic-is-not-really-a-bear.

- Disease Axes (6/19/2022):
  - Kinney GL, Santorico SA, **Young KA**, Cho MH, Castaldi P, San Jose Estapar R, Ross J, Dy J, Make B, Regan E, Lynch D, Everett DC, Lutz SM, Silverman EK, Washko G, Crapo J, Hokanson JE. Identification of Chronic Obstructive Pulmonary Disease Axes That Predict All-Cause Mortality: The COPDGene Study. *American Journal of Epidemiology.* 2018 Oct 1;187(10):2109-2116. doi.org/10.1093/aje/kwy087.
- Kruskal-Wallis Rank Sum Test (6/14/2022):
  - *Kruskal Wallis H test: Definition, examples, assumptions, SPSS*. Statistics How To. (2022, April 7). Retrieved June 14, 2022, from https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/kruskal-wallis/.
- Emphysema (6/15/2022):
  - *NCI Dictionary of Cancer terms*. National Cancer Institute. (n.d.). Retrieved June 15, 2022, from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/emphysema.
- GOLD Variable (6/19/2022):

- Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American journal of respiratory and critical care medicine. 2007;176(6):532-55. Epub 2007/05/18. doi: 10.1164/rccm.200703-456SO. PubMed PMID: 17507545.

- Airway Disease (6/15/2022):

  - *Reactive airway disease diagnosis & treatment*. UPMC. (n.d.). Retrieved June 15, 2022, from https://www.upmc.com/services/south-central-pa/allergy-asthma-immunology/asthma/reactive-airway-disease#:~:text=What%20is%20reactive%20airway%20disease,swell%2C%20and%20cause%20breathing%20problems.

- Obesity Facts (12/20/2021):

  - Ritchie, H., & Roser, M. (2017, August 11). *Obesity*. Our World in Data. Retrieved December 20, 2021, from https://ourworldindata.org/obesity.

- Standard Error of the Mean (SEM) (6/14/2022):

  - Tuovila, A. (2022, March 13). *Standard error of the mean vs. standard deviation: The difference*. Investopedia. Retrieved June 14, 2022, from https://www.investopedia.com/ask/answers/042415/what-difference-between-standard-error-means-and-standard-deviation.asp#:~:text=SEM%20is%20calculated%20by%20taking,root%20of%20the%20sample%20size.

- Final GOLD Variable (6/19/2022):

  - Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. Respiratory research. 2014;15:89. Epub 2014/08/07. doi: 10.1186/s12931-014-0089-y. PubMed PMID: 25096860; PubMed Central PMCID: PMC4256936.

- Bonferonni Corrections (6/14/2022):

  - {{r.PageAuthors}}. (n.d.). *What is the Bonferroni Correction?* AAOS. Retrieved June 14, 2022, from https://www.aaos.org/aaosnow/2012/apr/research/research7/#:~:text=The%20Bonferroni%20correction%20is%20an,number%20of%20comparisons%20being%20made.

- Subtypes of COPD Have Unique Distributions & Differential Risk of Mortality:

  - Young, K. A., Regan, E. A., Han, M. L. K., Lutz, S. M., Ragland, M., Castaldi, P. J., Washko, G. R., Cho, M. H., Strand, M., Curran-Everett, D., Beaty, T. H., Bowler, R. P., Wan, E. S., Lynch, D. A., Make, B. J., Silverman, E. K., Crapo, J. D., Hokanson, J. E., Kinney, G. L., & COPDGene® Investigators. (2019, November). *Subtypes of COPD have unique distributions and differential risk of mortality*. Chronic obstructive pulmonary diseases (Miami, Fla.). Retrieved January 20, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7020845/.

- Disease Axes (6/19/2022):

- **Young KA**, Strand M, Ragland MF, Kinney GL, Austin EE, Regan EA, Lowe KE, Make BJ, Silverman EK, Crapo JD, Hokanson JE; COPDGene® Investigators. Pulmonary Subtypes Exhibit Differential Global Initiative for Chronic Obstructive Lung Disease Spirometry Stage Progression: The COPDGene® Study. *Chronic Obstructive Pulmonary Disease*. 2019 Nov;6(5):414-429. doi: 10.15326/jcopdf.6.5.2019.0155. PubMed PMID: 31710796; PubMed Central PMCID: PMC7020848.