ENGR 298 - Data Analysis
Blood Oxygen Data Final Project

Thomas Wasylenko

Study

Background:

“Pulse oximetry is routinely used for monitoring patients' oxygen saturation levels with little regard to the variability of this physiological variable. There are few published studies on oxygen saturation variability (OSV), with none describing the variability and its pattern in a healthy adult population. The aim of this study was to characterize the pattern of OSV using several parameters: the regularity (sample entropy analysis), the self-similarity (detrended fluctuation analysis (DFA)), and the complexity (multiscale entropy (MSE) analysis). Secondly, to determine if there were any changes that occur with age.”

Procedure:

1. Clean the pulse oximeter and place on finger of participants choosing
2. Place the pulse pressure transducer on the adjacent finger
3. Wrap the Respiratory band around the umbilicus of the participant
4. Preferably have participant sitting with the fingers relatively still
5. Test equipment to ensure correct readings
6. Once the equipment has been checked stop the test and start the official recording

Parameters:

-Healthy Adult

-Gender

-Smoking Status

-BMI

-Age

Pulse Oximeter Picture

PhysioNet: Pattern Analysis of Oxygen Saturation Variability

Participant No	Gender	Smoking Status	BMI	Age
1	Female	Non-Smoker	20.5	23
2	Female	Non-Smoker	21.7	21
3	Female	Smoker	23.8	45
4	Female	Ex-Smoker	23.2	66
5	Male	Non-Smoker	25.1	70
6	Male	Non-Smoker	22	20
7	Female	Non-Smoker	19.8	19
8	Female	Non-Smoker	24.7	22
9	Female	Non-Smoker	20.7	20
10	Male	Non-Smoker	23.6	20
11	Female	Non-Smoker	28.2	41
12	Female	Non-Smoker	24.4	20
13	Female	Non-Smoker	24.1	45
14	Male	Non-Smoker	26.8	49
15	Male	Non-Smoker	23.2	22
16	Male	Non-Smoker	20.6	20
17	Femae	Non-Smoker	23.1	21
18	Male	Non-Smoker	26.2	42

Participant No	Gender	Smoking Status	BMI	Age
19	Female	Non-Smoker	24.3	62
20	Male	Non-Smoker	19.9	48
21	Male	Ex-Smoker	24.5	60
22	Female	Ex-Smoker	18.5	56
23	Female	Non-Smoker	24.2	24
24	Male	Non-Smoker	23.9	45
25	Female	Non-Smoker	19.8	38
26	Male	Smoker	24.6	21
27	Male	Non-Smoker	26.7	23
28	Female	Non-Smoker	18.9	21
29	Male	Smoker	26.8	23
30	Female	Ex-Smoker	17.9	21
31	Female	Non-Smoker	21.3	35
32	Male	Non-Smoker	23	20
33	Male	Ex-Smoker	28.4	55
34	Female	Non-Smoker	20.3	20
35	Male	Non-Smoker	26.5	42
36	M	Non-Smoker	21.7	20

Problem Definition

Take the data collected and perform analysis to see how and what physical factors have the greatest importance on a person's blood oxygen values.

Question: What physical factor has the greatest affect on blood oxygen values leading to the decrease in a person's levels?

Hypothesis:

-Gender will not play significant role

-Smoking status and BMI will be significant

Process

- Format blood oxygen participant data
- Perform 1-way Anova tests over all the participants
- Generate graphs and charts of the factors compared to each other Ex: (Male vs Female)
- Multiple comparison test to see groups of significance
- Perform N-way Anova tests to compare each physical factor

Format One-Way Anova

Editor - anova_bloodoxygen.m | Variables - tbld

anova_bloodoxygen.m | anova_test.m | mult_compare.m | test.m | anova_nway.m

```
1 -    filename = "reformat_osv_data.xlsx";
2
3 -    gender = xlsread(filename, "A2:A37");
4 -    smoking = xlsread(filename, "B2:B37");
5 -    bmi = xlsread(filename, "C2:C37");
6 -    age = xlsread(filename, "D2:D37");
7 -    data = xlsread(filename, "E2:E37");
8
9 -    save("pat_data.mat", 'gender', 'smoking', 'bmi', 'age', 'data');
10
```

Process:

- Format Data
  - Male 0, Female 1
  - Non-Smoker 0, Smoker/Ex-Smoker 1
- Load the data into Matlab
- Make each factor its own variable
- Perform Anova test to find significance in factors
- Perform Analysis of the data handed back

Gender	Smoking	BMI	Age	Average
1	0	20.5	23	98.239
1	0	21.7	21	98.163
1	1	23.8	45	94.412
1	1	23.2	66	96.344
0	0	25.1	70	93.748
0	0	22	20	96.468
1	0	19.8	19	97.955
1	0	24.7	22	98.435
1	0	20.7	20	98.703
0	0	23.6	20	97.873
1	0	28.2	41	98.801
1	0	24.4	20	97.364
1	0	24.1	45	98.009
0	0	26.8	49	98.185
0	0	23.2	22	96.805
0	0	20.6	20	98.606
1	0	23.1	21	99.001
0	0	26.2	42	98.184
1	0	24.3	62	98.704
0	0	19.9	48	98.504
0	1	24.5	60	98.2
1	1	18.5	56	96.524
1	0	24.2	24	96.976
0	0	23.9	45	96.275
1	0	19.8	38	96.87
0	1	24.6	21	96.523
0	0	26.7	23	97.77
1	0	18.9	21	98.33
0	1	26.8	23	98.186
1	1	17.9	21	99.324
1	0	21.3	35	99.508
0	0	23	20	98.645
0	1	28.4	55	97.126
1	0	20.3	20	98.838
0	0	26.5	42	97.585
0	0	21.7	20	98.218

Editor - anova_test.m | Variables - tbld

anova_bloodoxygen.m | anova_test.m | mult_compare.m | test.m | anova_nway.m

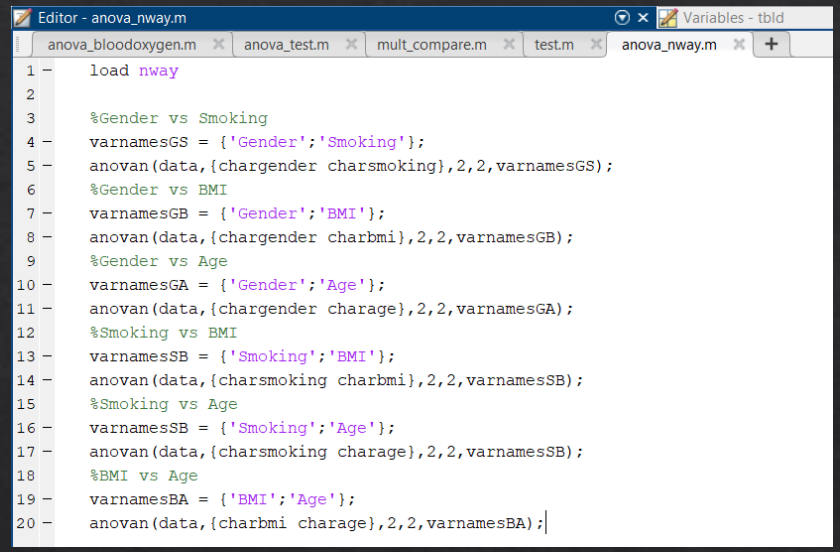
```
1 -    load pat_data
2
3 -    [p_gender,tbl_gender,stats_gender] = anova1(data, gender);
4 -    [p_smoking,tbl_smoking,stats_smoking] = anova1(data, smoking);
5 -    [p_bmi,tbl_bmi,stats_bmi] = anova1(data, bmi);
6 -    [p_age,tbl_age,stats_age] = anova1(data, age);
```

Format N-Way Anova

Gender	Smoking	BMI	Age	Average
Female	Non-Smoker	Healthy	Young	98.239
Female	Non-Smoker	Healthy	Young	98.163
Female	Smoker	Healthy	Mid	94.412
Female	Ex-Smoker	Healthy	Old	96.344
Male	Non-Smoker	Overweight	Old	93.748
Male	Non-Smoker	Healthy	Young	96.468
Female	Non-Smoker	Healthy	Young	97.955
Female	Non-Smoker	Healthy	Young	98.435
Female	Non-Smoker	Healthy	Young	98.703
Male	Non-Smoker	Healthy	Young	97.873
Female	Non-Smoker	Overweight	Mid	98.801
Female	Non-Smoker	Healthy	Young	97.364
Female	Non-Smoker	Healthy	Mid	98.009
Male	Non-Smoker	Overweight	Mid	98.185
Male	Non-Smoker	Healthy	Young	96.805
Male	Non-Smoker	Healthy	Young	98.606
Femae	Non-Smoker	Healthy	Young	99.001
Male	Non-Smoker	Overweight	Mid	98.184
Female	Non-Smoker	Healthy	Mid	98.704
Male	Non-Smoker	Healthy	Mid	98.504
Male	Ex-Smoker	Healthy	Mid	98.2
Female	Ex-Smoker	Underweight	Mid	96.524
Female	Non-Smoker	Healthy	Young	96.976
Male	Non-Smoker	Healthy	Mid	96.275
Female	Non-Smoker	Healthy	Young	96.87
Male	Smoker	Healthy	Young	96.523
Male	Non-Smoker	Overweight	Young	97.77
Female	Non-Smoker	Healthy	Young	98.33
Male	Smoker	Overweight	Young	98.186
Female	Ex-Smoker	Underweight	Young	99.324
Female	Non-Smoker	Healthy	Mid	99.508
Male	Non-Smoker	Healthy	Young	98.645
Male	Ex-Smoker	Overweight	Mid	97.126
Female	Non-Smoker	Healthy	Young	98.838
Male	Non-Smoker	Overweight	Mid	97.585
Male	Non-Smoker	Healthy	Young	98.218

Formatting:

- Gender
  - Male, Female
- Smoking Status
  - Non-Smoker, Ex-Smoker, Smoker
- BMI (Per BMI CDC)
  - <18.5 Underweight
  - 18.5-25 Healthy
  - 25-30 Overweight
  - >30 Obesity
- Age
  - 0-30 Young
  - 31-64 Mid
  - 65 above Old



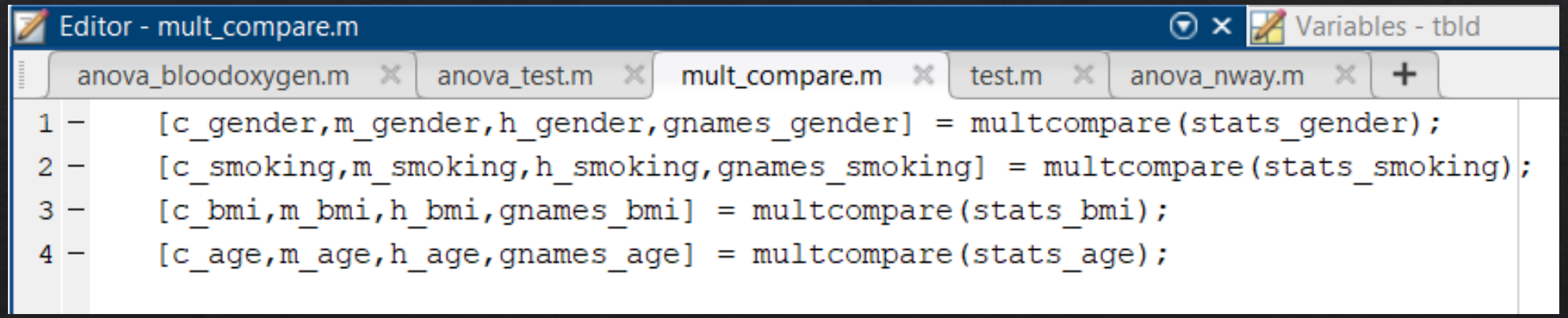
Process:

- Format data
- Load data as character arrays
- Perform Anova test comparing all combinations
- Perform analysis of data

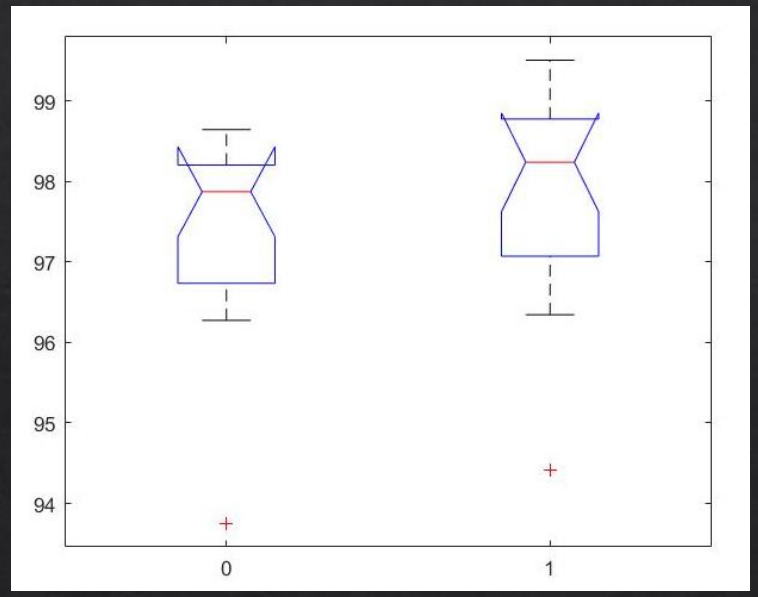
Multiple Comparison

Process:

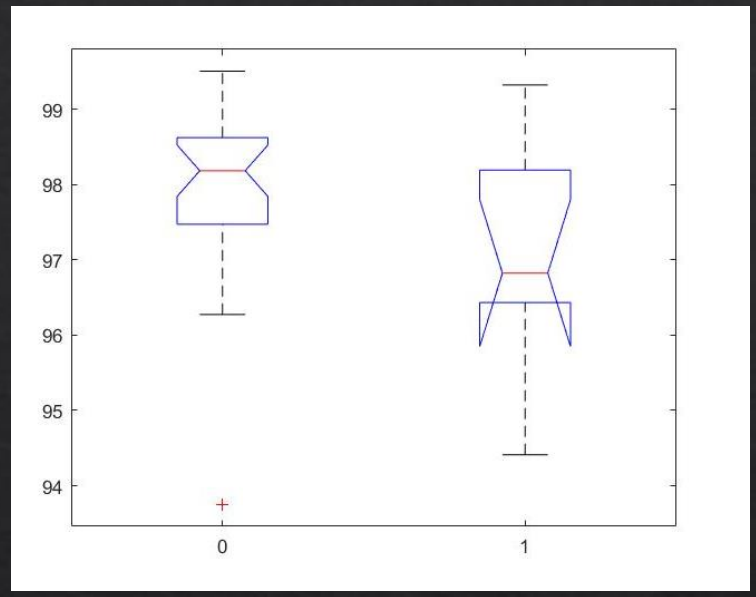
- Collect data from one-way Anova
- Preform Multiple Comparison for each factor
- Analyze data to find groups with large significance



Findings One-Way Anova



Gender Factor Test
Male 0, Female 1



Smoking Factor Test
Non-Smoker 0, Smoker 1

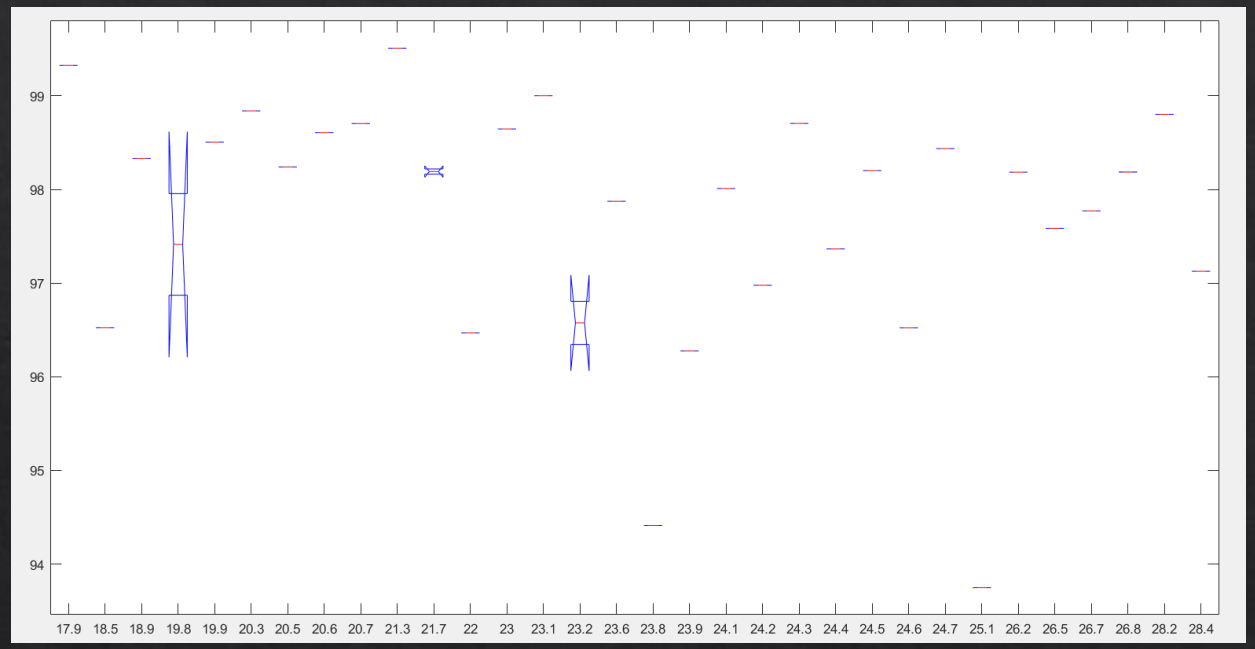
ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	1.868	1	1.868	1.21	0.2792
Error	52.527	34	1.54491		
Total	54.3951	35			

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	4.027	1	4.02697	2.72	0.1084
Error	50.3681	34	1.48141		
Total	54.3951	35			

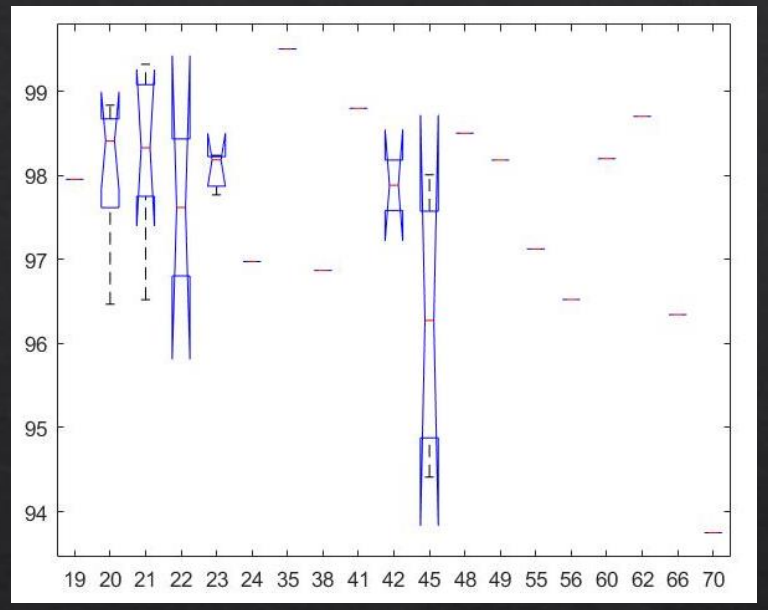
Findings One-Way Anova



BMI Factor Test

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	53.6987	31	1.73222	9.95	0.0185
Error	0.6964	4	0.1741		
Total	54.3951	35			

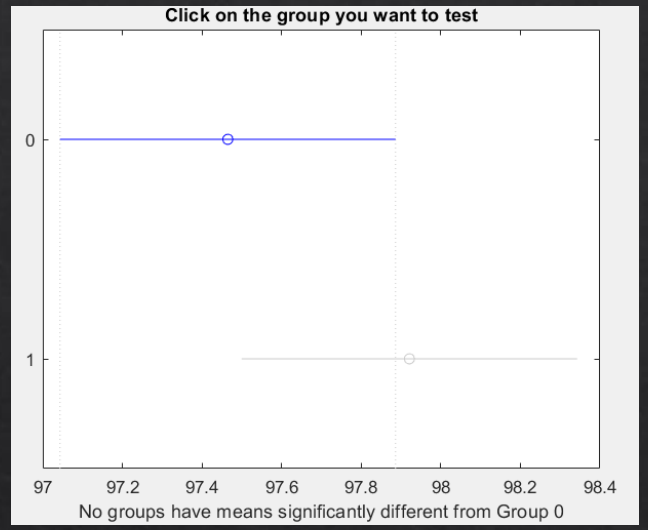


Age Factor Test

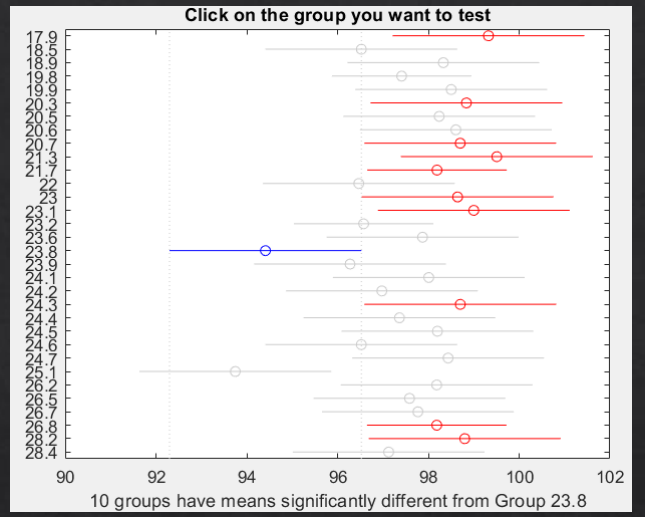
ANOVA Table

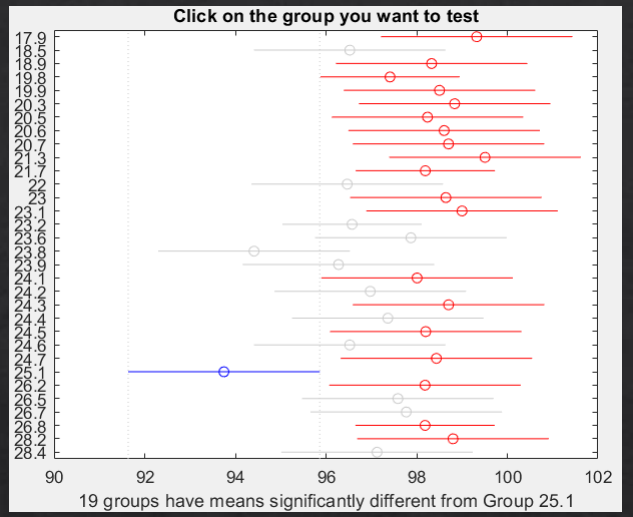
Source	SS	df	MS	F	Prob>F
Groups	36.84	18	2.04667	1.98	0.0826
Error	17.5551	17	1.03265		
Total	54.3951	35			

Findings Multiple Comparisons

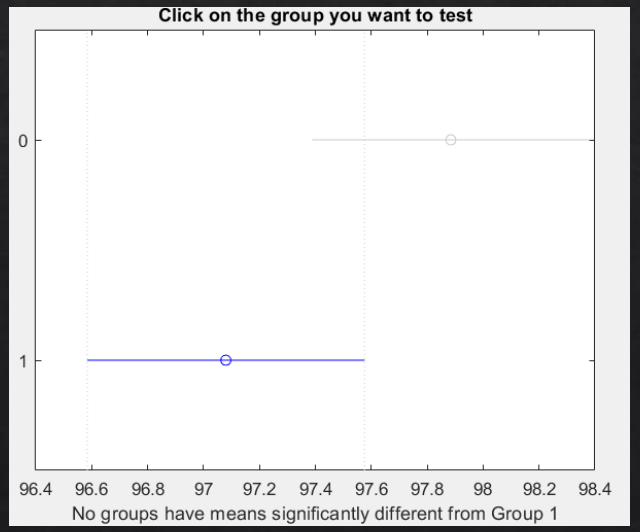


Gender Multi Test
Male 0, Female 1

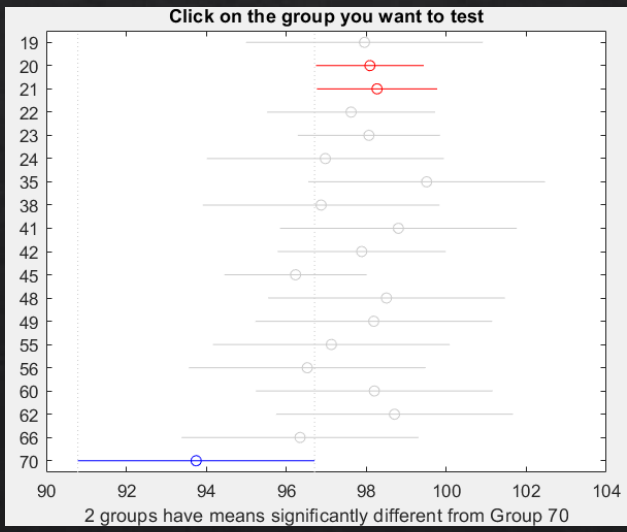




BMI Multi Test



Smoking Multi Test
Non-Smoking 0
Smoking 1



Age Multi Test

Findings N-Way Anova

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	Gender	2.2121	2	1.10604	0.87	0.4306
	Smoking	5.5339	2	2.76694	2.17	0.1324
#	Gender*Smoking	8.7859	2	4.39296	3.45	0.0454
	Error	36.9762	29	1.27504		
	Total	54.3951	35			

Gender vs Smoking
Smoking has greater significance

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	Gender	2.4348	2	1.21741	0.73	0.4887
	BMI	0.0627	2	0.03136	0.02	0.9813
#	Gender*BMI	1.4288	1	1.42876	0.86	0.361
	Error	49.8045	30	1.66015		
	Total	54.3951	35			

Gender vs BMI
Gender has slightly greater significance

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	Gender	2.3269	2	1.16347	1.01	0.3765
	Age	14.8038	2	7.40188	6.43	0.0049
#	Gender*Age	3.1022	2	1.55111	1.35	0.2758
	Error	33.39	29	1.15138		
	Total	54.3951	35			

Gender vs Age
Age has greater significance

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	Smoking	6.5703	2	3.28516	2.35	0.113
	BMI	0.8764	2	0.43821	0.31	0.7331
#	Smoking*BMI	6.6052	2	3.30262	2.37	0.1118
	Error	40.4925	29	1.39629		
	Total	54.3951	35			

Smoking vs BMI
Smoking has greater significance

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	Smoking	7.6392	2	3.81961	5.54	0.0094
	Age	16.7943	2	8.39713	12.17	0.0002
#	Smoking*Age	11.8598	3	3.95327	5.73	0.0035
	Error	19.3201	28	0.69		
	Total	54.3951	35			

Smoking vs Age
Age has greater significance

Analysis of Variance

	Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
	BMI	0.0255	2	0.01274	0.01	0.9888
	Age	14.8744	2	7.4372	6.56	0.0046
#	BMI*Age	7.0695	3	2.3565	2.08	0.1255
	Error	31.7242	28	1.13301		
	Total	54.3951	35			

BMI vs Age
Age has greater significance

Summary

Initial Question: What physical factor has the greatest affect on blood oxygen values leading to the decrease in a person's levels?

Hypothesis:

- -Gender will not play significant role
- -Smoking status and BMI will be significant

GitHub Link

After Analysis:

- Age has the greater significance over all other factors
- Smoking is the next significant factor
- BMI is least significant

Final Answer in relation to Hypothesis:

My hypothesis was not proven to be correct entirely, only that smoking status plays a significant role.

Challenges:

- Small Sample Size for proper analysis
  - Possible Solution: Larger Sample Size with greater variability in factors