Biostatistics - Dr. Patrick
BMEN 350; Section 201
Saurabh Dhole 626 002 135
October 2nd, 2021

Project 2

A. Read and Reflect

Upon reviewing the esteemed article called, "Statistical tests, P values, confidence intervals, and power: A guide to misinterpretation" by Dr. Greenland, I was intrigued to find out about a key common misconception about P values. I reflect on my new found knowledge from Dr. Greenland's article here. Dr. Greenland has made it clear in the article that statistical tests have been misrepresented and abused for decades. This misrepresentation and abuse of statistical tests still runs rampant today. One of the main reasons for this misrepresentation and abuse of statistical tests of significance is a key misconception about the P value. This misconception is as follows: The P value is the probability that the test hypothesis is true. This is not the case, as the P value already assumes that the hypothesis is true. I observed that the P value merely establishes the degree to which the data conform to the pattern predicted by the test hypothesis. I found this very interesting because I myself have been guilty of thinking that the P value is the probability that the test hypothesis is true. Upon reviewing this information about the misconceptions of P values and upon learning about P values in the BMEN 350 class, I vow to not adhere to such a misconception ever again. As a Biomedical Engineer, this information is very important to me as I am currently accumulating 10 industry level statistical significance tests into my tool belt. When I wield one of these statistical significance tests in a clinical or corporate setting, I must ensure that I do not abuse or allow the misinterpretation of how the statistical significance test were used in my studies. This information about the misconception about the P value is indeed very important to me because I was under the wrong impression regarding P values until I enrolled in BMEN 350 and until I reviewed this article. I plan on using this information when I am performing data analysis in my undergraduate research projects.

After thoroughly reviewing Dr. Altman's article called "Interpreting P values", I was elated upon learning about some cautionary advice on P values. I would like to reflect on my findings from Dr. Altman's article herewith. In the article, Dr. Altman clearly states that the P value must be applied with caution as it can occasionally lead one astray. Dr. Altman explains that the proper use of P values requires that they be properly computed. I observed that Dr. Altman emphasizes the importance of paying appropriate attention to the sampling design when properly computing P values. When performing an experimental study, the interpretation of the P value can be assisted by including heuristics such as the Bayes factor. As a Biomedical Engineer, this information is crucial to my application of the P value in my studies, career, and research. I say this because in a research, clinical, or corporate setting, I will have to ensure that the P value is interpreted by audiences properly. To ensure the proper interpretation of the P value, I must pay close attention to the sample design, and I must consider whether I should bolster the interpretation of the P value with the Bayes factor. In a general sense, this article is indeed very important to me as it has aided me in my understanding of the P value. Now that I am equipped with this knowledge from Dr. Altman's article, I will be careful about my interpretation of P values when I am presented with studies. I will ask myself questions such as, "should the authors of this study have paid more attention to their sampling design?" instead of just taking the information from the P value at face value. I plan on applying my new found information about the P value to my undergraduate research studies and to my projects in the BMEN 350 course.

Dr. Halsey's eloquently written article called "The fickle P value generated irreproducible results", certainly enlightened me on some of the difficulties faced when relying on the P value. Here I present some of the reflections that I had upon reviewing this great article. Dr. Halsey has argued in her article that the reliability and reproducibility of scientific experiments are taking some heat. Dr. Halsey states that there is rampant irreproducibility within the scientific community. Further, Dr. Halsey has provided a reason for this lack of repeatability. This reason is the wide sample-to-sample variability. I observed that Dr. Halsey recommends that analyses should not be solely made just based off of the P value. I learned about a concept called exaggerated effect sizes. This concept essentially states that a significance test performed with a small sample size, can indeed result in a lower P value. This is because the smaller samples are less representative. To ameliorate such discrepancies, the estimated effect size must also be reported along with P values. This measure, Dr. Halsey believes, will aid in reducing some of the irreproducibility of scientific experiments. As a Biomedical Engineer, this information is

indeed very important for me. I say this because in quality engineering internships and in research, I will have to apply the P value to test for statistical significance. I must not get too comfortable using the P value, as the sample size has effects on the P value. In a general sense, this information is crucial for me as it will remind me to analyze the effect size of a sample before I report P values to the Principal Investigator. I plan to apply Dr. Halsey's cautionary advice on the P value to projects in the BMEN 350 course and to projects in quality assessment internships as well.

Following my thorough reading of Dr. Krzywinski's informative article called "Comparing Samples-Part I", my understanding of the t-test was significantly reinforced. I would like to reflect on my interpretation of Dr. Krzywinski's article here. Dr. Krzywinski explained that a paired t-test is applicable when we have matched-sample experiments. I had learned about this topic in the textbook required for BMEN 350, and I was intrigued when I saw this concept reappear in Dr. Krzywinski's article. Another key point that Dr. Krzywinski discusses in his article, that I also had learned about in the textbook is the fact that the t-test is the optimal method for comparing means, given that the assumptions are fulfilled for the t-test. This is where there may be some problems in the validity of the t-test. I observed that this is because the t-test assumes that the samples are drawn from populations that have normal distribution. I conjectured that this may indeed lead to problems when relying on the t-test because it is not certain that the population is of normal distribution. As a Biomedical Engineer, this information is crucial to my application of the t-test to my undergraduate research, and undergraduate internships. I say this because there will be multiple occasions on which I will have to use the t-test to ensure statistical significance in internships. When I perform such t-tests, I must ensure that I keep in mind some of the assumptions of the t-test so that I can accurately present the results to my supervisors. In a general sense, this information discussed by Dr. Krzywinski is essential to me because it has reinforced my understanding of the t-test and some of its limitations. I plan to apply this knowledge to my BMEN 350 course projects as well as my undergraduate research projects.

Following my review of Dr. Krzywinski's informative article called "Comparing Samples-Part II", I was thoroughly enlightened on a concept that I had initially not known about. I would like to reflect upon my new found knowledge here. Before reading this article, I was under the impression that the P value is only affected by smaller sample sizes, as was discussed in Dr. Halsey's article. However, after reading this article by Dr. Krzywinski, my understanding of how the P value changes based on the number of experimental results has been warped. Dr. Krzywinski states that "the use of P values, which assign a measure of rarity to a single experimental outcome, is misleading when many experiments are considered". Therefore, these values need to be adjusted and reinterpreted according to Dr. Krzywinski. Dr. Krzywinski explains that there are methods to perform this readjustment or reinterpretation. These methods are called multiple-testing corrections. I observed that Dr. Krzywinski has outlined a few of these multiple-testing correction in his article. As a Biomedical Engineer, this information is indeed very important for me. I say this because during my quality engineering internships I will most likely have to perform significance tests on a data set containing a large number of experimental results. In order for my supervisors to be able to properly understand the story being told by the experimental results, I will have to apply some of these multiple-testing corrections to the P values obtained. In a general sense, this article is essential for me as my initial incorrect impressions of the effects of large numbers of experiments on P value have been corrected. I plan on applying the multiple-testing correction methods to my undergraduate research in Lymphatics, as I think they may aid my Principal Investigator in seeing trends of drug effectiveness against head and neck tumors.

B. Problem 1

We are asked to determine whether or not there is a detectable difference in performance on the walking test between patients diagnosed with COPD and patients in the control group. The performance on the waking test is measured in meters. Let the null hypothesis (H_o) be established as follows: there is no difference between the mean distance walked (in meters) between the COPD patients and the patients in the control group. Let the alternative hypothesis (H_1) be established as follows: There is a difference in the mean distance walked (in meters) between the COPD patients and the patients in the control group. The sample size of the COPD group is equal to the sample size of the control group. The variance of the COPD group was determined to be 4220.59. The variance of the control group was determined to be 4220.2. Since these variances are very similar, it was assumed that the variances of the two samples were the same. Hence, a 2-sample t-test of equal size and variance was performed. This 2-sample t-test was performed in Microsoft excel. The alpha was set to 0.05. The P value was determined to be 7.29E-10. It can be observed that 7.29E-10 is sufficiently smaller than 0.05 or in other words $p < \alpha$. From this observation, we must reject the null hypothesis! We can indeed state that there is a detectable difference in performance on the walking test between COPD patients and patients in the control group. Figures 1 and 2 below depict the data of distance walked in bar graph and in box plot format. I prefer the bar graph because it easily allows one to observe that there is a detectable difference in mean distance walked between the two groups.

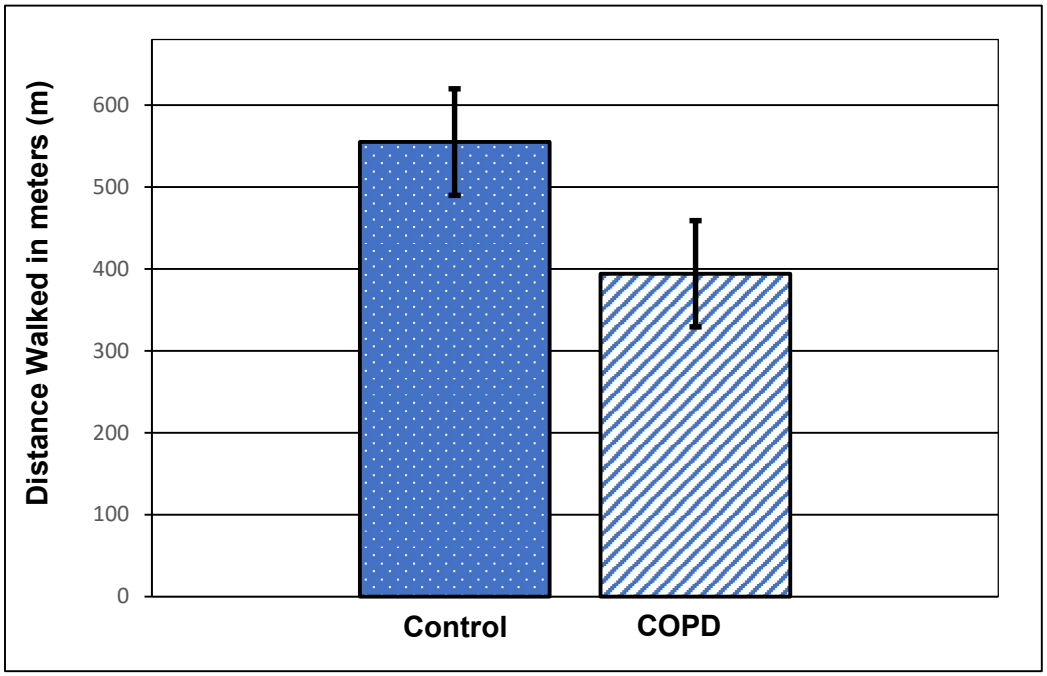


Figure 1: The mean distance walked in meters for control and COPD patients as represented as a bar chart. The mean distance walked for the control patients was 555 ± 64.96 meters (mean ± std.dev.). The mean distance walked for the COPD patients was 394.09 ± 64.96 meters (mean ± std.dev.). It can be seen that there is indeed a detectable difference between the mean distance walked of the control patients and the mean distance walked of the COPD patients. This is because the standard error bars do not overlap. The fact that these error bars do not overlap indicates that there is indeed a detectable difference in the mean distance walked between the COPD and control patient groups.

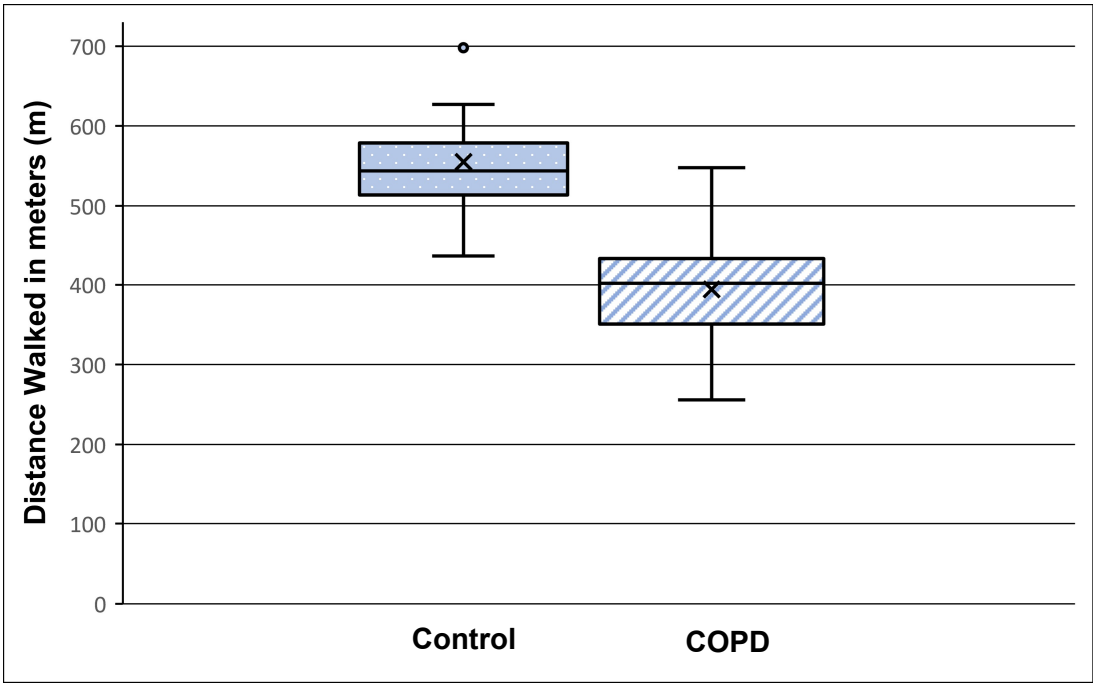


Figure 2: The distance walked in meters for control and COPD patients as represented as a box plot. The 25th, 50th, and 75th percentiles of distance walked in meters for control patients are 515, 544, and 572 meters respectively. With a minimum and maximum of 436 and 700 meters respectively. The 25th, 50th, and 75th percentiles of distance walked in meters for COPD patients are 358, 402, and 428 meters respectively. With a minimum and maximum of 256 and 548 meters respectively

C. Problem 2

We are asked to test the hypothesis that the mean degree of itching is the same for eyes treated with active drug vs. eyes treated with placebo. This hypothesis was tested using a t-test and an ANOVA. The degree of itching is measured via a scale from 1 to 4 (1=none, 2=mild, 3=moderate, 4=severe). The null hypothesis (H_o) was established to be as follows: There is no difference in the mean degree of itching between eyes treated with active drug and eyes treated with placebo. The alternative hypothesis (H_1) was established to be as follows: There is a difference in the mean degree of itching between eyes treated with active drug and eyes treated with placebo. The alpha was set to 0.05. The variance of the itch scores of the eyes treated with active drug was determined to be 1.28. The variance of the itch scores of the eyes treated with placebo was determined to be 0.62. The sample size of the group treated with active drug was equal to the sample size of the group treated with placebo. Hence, a 2-tailed t-test of unequal variance and equal sample size was performed. The P value derived from the t-test was 0.036. It can be observed that 0.036 is sufficiently less than 0.05 or in other words $p < \alpha$. From this observation, we must reject the null hypothesis! We have reason to state that there is a detectable difference in the mean degree of itching between eyes treated with active drug and eyes treated with placebo. From this rejection of the null hypothesis, we can conjecture that the difference in the mean degree of itching between the two groups may indeed have been brought about by the presence of the active drug. This is assuming that there were no confounding variables at play in this experiment, and that the only difference between the two samples was the presence of the active drug or the placebo. If these assumptions are in fact true, then it may be safe to state that the improvement in mean degree of itching between the placebo group and the active drug group may have been caused by the drug. The null hypothesis was also tested using a single factor ANOVA. The alpha value was set at 0.05. The P value obtained from the ANOVA was determined to be 0.034. It can be clearly observed that 0.034 is sufficiently less than 0.05 or in other words $p < \alpha$. Following this observation, we must reject the null hypothesis! It can be stated that there is a detectable difference in the mean degree of itching between eyes treated with active drug and eyes treated with placebo. The P value obtained from the t-test and the P value obtained from the ANOVA are very similar, and both of these P values allowed us to reject the null hypothesis (H_o) that there is no difference in the mean degree of itching between eyes treated with active drug and eyes treated with placebo. The results of the ANOVA also point towards the fact that the drug may have resulted in the improvement in mean degree of itching between the placebo group and the active drug group (given that there were no confounding variables at play, and given that the only difference between the two samples was the presence of the active drug or the placebo).

D. Problem 3

We are initially tasked with just determining if there is at least one difference or inequality between the means of the high NAPAP, low NAPAP, and control groups in the 1968 batch. The data in the 1968 batch was split up according to the three groups: high NAPAP, low NAPAP, and control. A single variable ANOVA was performed. The null hypothesis (H_o) was established to be as follows: There are no differences between the means of the high NAPAP, low NAPAP, and control groups. The alternative hypothesis (H_1) was established to be as follows: There is at least one difference between the means of the high NAPAP, low NAPAP, and control groups. The alpha was set to 0.05. The P value from the ANOVA was determined to be 0.000183. It can be clearly observed that 0.000183 is sufficiently less than 0.05, or in other words $p < \alpha$. From this observation, we must reject the null hypothesis. It is safe to say that there is at least one difference between the means of the high NAPAP, low NAPAP, and control groups. How can we find out which groups are indeed different? We perform the Tukey-Krammer post hoc test. This test clearly showed that there is a significant difference of the means of the high NAPAP and low NAPAP groups. This is because the absolute difference between the means of the High NAPAP group and the low NAPAP group is greater than the critical range (0.157>0.109). The Tukey-Krammer post hoc also clearly showed that there is a significant difference between the means of the high NAPAP and control groups. This is because the absolute difference between the High NAPAP and control group is greater than the critical range (0.179>0.109). The Tukey-Krammer post hoc further shows that there is no significant difference between the means of the low NAPAP and control groups. This is because the absolute difference between the low NAPAP and control groups is less than the critical range (0.022 < 0.109). We were then tasked with determining if there is at least one difference or inequality between the means of the control groups throughout the years 1968, 1969, 1970, 1971, 1972, 1975, and 1978. The null hypothesis (H_o) in this case was established as follows: There are no differences or inequalities

between the means of the control groups throughout the years 1968, 1969, 1970, 1971, 1972, 1975, and 1978. The alternative hypothesis (H_1) in this case was established to be as follows: There is at least one difference or inequality between the means of the control groups throughout the years 1968, 1969, 1970, 1971, 1972, 1975, and 1978. It was determined via a single factor ANOVA that there was indeed at least one difference or inequality between the means of the control groups in the years 1968, 1969, 1970, 1971, 1972, 1975, and 1978. This is because the P value resulting from this ANOVA was sufficiently less than the alpha of 0.05 (5.67E-5 < 0.05). Thus, the null hypothesis stating that there was no difference or inequality between the control groups throughout the above listed years was rejected! Thus, we do have reason to believe that there is at least one detectable difference or inequality between the means of the control groups in the above listed years. We then needed to find out exactly which year's control group mean was different from the control group mean of the baseline (year 1968). To find this out, a Dunnett post hoc was performed. Based on the results of this Dunnett post hoc, it was determined that the mean of the control group of the year 1970 was different than the mean of the control group of the year 1968 (baseline). This is because the absolute difference between the means of the control group in the year 1970 and the control group in the baseline year (1968), was greater than the critical range (0.084 > 0.065). Based off of this Dunnett post hoc test, we were able to conclude that the mean of the control group in 1970 was detectably different than the mean of the control group of the year 1968.

E. Problem 4

Contingency table for Smaller Sample	Observed		
	High Salt	Normal	Row Total
Died of CVD	5 (4.08)	30 (30.92)	35
Died of other	2 (2.92)	23 (22.08)	25
Column Total	7	53	60

Figure 3: This is the contingency table for the smaller sample. A total of 35 have died of CVD, 5 out of these 35 were on a high salt diet. A total of 25 have died of other causes, 2 out of these 25 were on a high salt diet. Inside the parentheses next to the observed values are the expected frequencies!

Contingency table for Larger Sample	Observed		
	High Salt	Normal	Row Total
Died of CVD	23 (19.49)	1977 (1980.51)	2000
Died of other	15 (18.51)	1885 (1881.48)	1900
Column Total	38	3862	3900

Figure 4: This is the contingency table for the larger sample. A total of 2000 have died of CVD, 23 out of these 2000 were on a high salt diet. A total of 1900 have died of other causes, 15 out of these 1900 were on a high salt diet. Inside the parentheses next to the observed values are the expected frequencies!

The contingency tables for the two scenarios given are displayed in Figures 3 and 4. There are certain circumstances under which Chi squared test, Chi squared with Yates correction, and Fisher's exact test should be conducted. I will discuss these circumstances here. The Chi squared test approximates a x^2distribution, and it works really well with large samples. The requirements needed to perform a Chi squared test are as follows: no cells with expected counts near 0, and no more than 20% of the cells should have expected counts less than 5. At least 80% of the cells must have expected frequencies that are greater than 5 in order for the Chi squared test to be deemed appropriate (expected frequencies can be found in the Appendix). For these reasons, the Chi squared test would not be appropriate for the first contingency table, Figure 3, as 50% of the cells have an expected frequency of less than 5. The conditions to perform a Chi squared test are not met by this first contingency table (Figure 3) so the Chi squared test is not appropriate here. However, the Chi squared test would be perfect for the contingency table

in Figure 4 as there are no cells with expected counts less than 5. The Chi squared with Yates correction is used to obtain values of the test statistic from much smaller samples. More specifically, the Chi squared with Yates correction is used exclusively for 2x2 contingency tables. For these reasons, the Chi squared with Yates correction is appropriate to use on the contingency table in Figure 3. The Chi squared with Yates correction would not be appropriate for the contingency table in Figure 4, as it is a much larger sample. The Fisher's exact test computes probability directly, and does not rely on approximation of x^2 distribution. The Fisher's exact test is used to obtain values of the test statistic from small samples, such as 2x2 tables. More specifically, the Fisher's exact test is used when there are cells with expected frequencies less than 5. This makes the Fisher's exact test perfect for the contingency table in Figure 3, as there are indeed cells that have an expected frequency that is smaller than 5. The Fisher's exact test may not be appropriate for the contingency table in Figure 4 as there are no cells with expected frequencies smaller than 5, also the contingency table in Figure 4 contains a much larger sample. In summary, Chi squared test would be appropriate for the contingency table in Figure 4, Chi squared with Yates correction would be appropriate for the contingency table in Figure 3, and the Fisher's exact test would be appropriate for the contingency table in Figure 3 as well. These three tests were used to obtain P values in both of the contingency tables. The results are as follows:

For the contingency tables in Figure 3 and 4, the null hypothesis (H_o) was established to be as follows: There is no association between diet and cause of death. The alternative hypothesis (H_1) was established to be as follows: There is an association between diet and cause of death. The alpha was set to 0.05, and the critical value was determined to be 3.84. Please view the results of Chi squared, Chi squared with Yates correction, and Fisher's exact tests in Table 1 below:

Table 1: Summary of P values obtained, via various methods, for contingency tables in Figures 3 and 4.

P Values obtained for the Contingency table in Figure 3 (smaller sample)			
Method	***P value***	***Chi-square statistic***	***Verdict***
Chi squared test	0.454	0.559	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis
Chi squared with Yates correction	0.733	0.115	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis
Fisher's exact test	0.688	N/A	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis
P Values obtained for the Contingency table in Figure 4 (larger sample)			
Method	***P value***	***Chi-square statistic***	***Verdict***
Chi squared test	0.252	1.312	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis
Chi squared with Yates correction	0.325	0.965	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis
Fisher's exact test	0.2595	N/A	P > alpha ; Chi squared statistic < critical value so, fail to reject the null hypothesis

It can be clearly deciphered from Table 1 that the P values obtained for the contingency table in Figure 3 (smaller sample) are all greater than the alpha of 0.05, regardless of the type of test used. Following this observation, we make the decision to fail to reject the null hypothesis that there is no association between diet and cause of death. It can also be clearly deciphered from Table 1 that the P values obtained for the contingency table in figure 4 (larger sample) are all greater than the alpha of 0.05, regardless of the type of test used. Following this observation, we make the decision to fail to reject the null hypothesis that there is no association between diet and cause of death. Comparisons between the Chi squared statistic and the critical value also support this decision. According to these above conclusions, and given that there are no confounding variables affecting these data, it is safe to state that there may not be an association between diet and cause of death. It must be observed from Table 1 that the P value differs between each method used. For example, the P value obtained from the Chi squared test is different from the P value obtained from the Fisher's exact test. Different P values result from using different tests because each of these different tests have their own unique formula. It must also be observed from Table 1 that the P values obtained for the contingency table in Figure 4 (lager sample) are smaller than the P values obtained for the contingency table in Figure 3 (smaller sample). The reasons for the smaller P value in the larger sample are as follows. Generally, if the sample size is larger, it is more likely that a study will find a significant association if one exists. In addition, the overall variability is decreased in a larger sample. As the sample size increases, the impact of random error is also reduced. This improved precision allows us to detect smaller and smaller differences and inequalities between groups. These are the reasons as to why the P values are observed to be lower for the larger sample when compared to the P values of the smaller sample.

F. Appendix

Problem 1: Microsoft excel was primarily used to perform the 2-sample t-test of equal size and variance.

	Distance Walked (m)	
	Control	COPD
1	619	283
2	512	402
3	523	407
4	586	402
5	436	340
6	515	445
7	562	548
8	544	344
9	531	358
10	534	419
11	572	393
12	541	469
13	551	393
14	492	420
15	698	463
16	700	438
17	571	428
18	502	364
19	557	336
20	482	256
21	627	368

Ho: The means are the same		Variance of Control =	4220.2
H1: The means are different		Variance of COPD =	4220.59
The variance seems to be the same.			
The sample size is the same.			
So let's perform a 2 sample t-test with equal size and variance			
Find sp:	64.96457		
Degrees of Freedom = 2n-2		40	
Find the t statistic:	8.025775076		
P value:	7.29422E-10		
Would reject the Null hypothesis!			
plot a bar graph, and plot a box plots.			

Problem 2: Microsoft excel was used to perform: (1) 2-tailed t-test of equal size and unequal variance, and a (2) one variable ANOVA. Displayed in that order.

Randomized Assignment

Subject	Eye	
	Left	Right
1	A	P
2	P	A
3	A	P
4	A	P
5	P	A
6	A	P
7	A	P
8	P	A
9	A	P
10	A	P

A= active drug P=placebo

Itching Scores

Subject	Eye	
	Left	Right
1	1	2
2	3	3
3	4	3
4	2	4
5	4	1
6	2	3
7	2	4
8	3	2
9	4	4
10	1	2

Subject	Drugged eye	Placebo eye
1	1	2
2	3	3
3	4	3
4	2	4
5	1	4
6	2	3
7	2	4
8	2	3
9	4	4
10	1	2
Variance:	1.28888889	0.622222222

2 tailed t-test, unequal variance, equal sample size.

	P Value:
Perform T test:	0.0360797

We must reject the null hypothesis

Ho: There is no difference between the mean itch score of the drugged and placebo groups.
H1: There is a difference between the mean itch score of the drugged and placebo groups.

(1)

Randomized Assignment

Subject	Eye	
	Left	Right
1	A	P
2	P	A
3	A	P
4	A	P
5	P	A
6	A	P
7	A	P
8	P	A
9	A	P
10	A	P

A= active drug P=placebo

Itching Scores

Subject	Eye	
	Left	Right
1	1	2
2	3	3
3	4	3
4	2	4
5	4	1
6	2	3
7	2	4
8	3	2
9	4	4
10	1	2

Subject	Drugged eye	Placebo eye
1	1	2
2	3	3
3	4	3
4	2	4
5	1	4
6	2	3
7	2	4
8	2	3
9	4	4
10	1	2
Variance:	1.28888889	0.622222222

We must reject the null hypothesis

Perform a single factor ANOVA

Drugged	Placebo
1	2
3	3
4	3
2	4
1	4
2	3
2	4
2	3
4	4
1	2

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Drugged	10	22	2.2	1.28889
Placebo	10	32	3.2	0.62222

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5	1	5	5.23256	0.03449	4.41387
Within Groups	17.2	18	0.95556			
Total	22.2	19				

(2)

Problem 3: Microsoft Excel was used to perform (1) ANOVA, (2) ANOVA+ Tukey-Krammer post hoc, and (3) ANNOVA+ Dunnett post hoc. Displayed in that order.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Level 1	100	107.31	1.0731	0.219399
Level 2	100	91.55	0.9155	0.054904
Level 3	100	89.34	0.8934	0.050809

ANOVA

ce of Varia	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between G	1.920609	2	0.960304	8.861302	0.000183	3.026153
Within Grc	32.18606	297	0.108371			
Total	34.10667	299				

(1)

High NAPAP	Low NAPAP	control
Level 1	Level 2	Level 3
0.97	0.7	1.48
0.88	0.84	0.78
0.96	0.47	0.79
0.74	0.81	0.8
0.85	0.75	0.84
1.06	0.76	0.94
0.65	0.91	0.86
0.65	0.87	1.5
0.94	0.99	0.8
0.86	0.85	0.88
0.78	0.66	0.78
0.84	0.85	1.2
0.75	0.77	1
0.82	0.74	0.72
0.76	0.89	0.76
2.1	0.9	0.66
1.38	0.61	0.71
1.8	0.75	0.75
0.77	1.04	0.72
0.76	0.8	1.19
1.2	0.66	0.9
0.82	0.75	0.99
0.99	0.72	0.6

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Level 1	100	107.31	1.0731	0.219399
Level 2	100	91.55	0.9155	0.054904
Level 3	100	89.34	0.8934	0.050809

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.920608667	2	0.960304333	8.861302	0.000183	3.026153
Within Groups	32.186058	297	0.108370566			
Total	34.10666667	299				

This shows that there is at least one difference amongst the means of the three groups. Now we just have to find out which ones are different. We do this via Tukey-Krammer

3.335

Tukey - Krammer Procedure		**Num df**	3	**Den df**	297
Comparison	**Absolute Difference**	**Critical Range**	**Results**		
1 to 2	0.1576	0.109787149	1!=2		
1 to 3	0.1797	0.109787149	1!=3		
2 to 3	0.0221	0.109787149	(2 not sig. diff. than 3)		

(2)

69 to 68	70 to control				78 to 68
0.0052	0.084	0.0289	0.0361	0.0031	0.0269
0.065285246	0.065285246	0.065285246	0.065285246	0.065285	0.065285
-0.060085246	0.018714754	-0.036385246	-0.02918525	-0.06219	-0.03839

Dunnett testing !

Comparisson	AbsoluteDiff.	CriticalRange	Result
69 to 68	0.0052	0.065285246	means not sig. diff.
70 to 68	0.084	0.065285246	means sig.diff.
71 to 68	0.0289	0.065285246	means not sig. diff.
72 to 68	0.0361	0.065285246	means not sig. diff.
75 to 68	0.0031	0.065285246	means not sig. diff.
78 to 68	0.0269	0.065285246	means not sig. diff.

Anova: Single Factor — Control with groups of year

0

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Year 1968	100	89.34	0.8934	0.050809
year 1969	100	88.82	0.8882	0.081007
year 1970	100	80.94	0.8094	0.017529
year 1971	100	86.45	0.8645	0.020902
year 1972	100	92.95	0.9295	0.017011
year 1975	100	89.03	0.8903	0.01768
year 1978	100	92.03	0.9203	0.018306

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.947079	6	0.157846	4.949438	5.67E-05	2.111645
Within Groups	22.10102	693	0.031892			
Total	23.04809	699				

Dunnett number

7 groups 693 dof and alpha 0.05

2.585

$$|\bar{x}_j - \bar{x}_0| > t_d\sqrt{2MS_W/n}$$

(3)

Problem 4: (1) Microsoft Excel was used to find the Chi squared test statistics values from the contingency tables. (2) A website called Social Science Statistics was used to find the P values of the various Chi squared tests (https://www.socscistatistics.com/tests/chisquare/default2.aspx). (3) Matlab was used to perform the Fisher's exact test.

Contingency table for Smaller Sample	Observed		
	High Salt	Normal	Row Total
Died of CVD	5	30	35
Died of other	2	23	25
Column Total	7	53	60

	Expected		
	High Salt	Normal	Row Total
Died of CVD	4.083333333	30.91666667	35
Died of other	2.916666667	22.08333333	25
Column Total	7	53	60

Chi Square	High Salt	Normal	Row Total
Died of CVD	0.205782313	0.027178796	0.232961109
Died of other	0.288095238	0.038050314	0.326145553
Column Total	0.493877551	0.065229111	0.559106662

CV	3.841459
P value	0.905727

Contingency table for Larger Sample			
Observed	High Salt	Normal	Row Total
Died of CVD	23	1977	2000
Died of other	15	1885	1900
Column Total	38	3862	3900

Expected	High Salt	Normal	Row Total
Died of CVD	19.48717949	1980.512821	2000
Died of other	18.51282051	1881.487179	1900
Column Total	38	3862	3900

Chi Square	High Salt	Normal	Row Total
Died of CVD	0.633232119	0.006230663	0.639462782
Died of other	0.666560125	0.006558593	0.673118718
Column Total	1.299792244	0.012789256	1.312581499

CV	3.841459
P value	0.726148

(1)

Contingency table for Smaller Sample	Observed		
	High Salt	Normal	Row Total
Died of CVD	5	30	35
Died of other	2	23	25
Column Total	7	53	60

	Expected		
	High Salt	Normal	Row Total
Died of CVD	4.083333333	30.91666667	35
Died of other	2.916666667	22.08333333	25
Column Total	7	53	60

Chi Square w/ Yates	High Salt	Normal	Row Total
Died of CVD	0.042517007	0.005615454	0.048132461
Died of other	0.05952381	0.007861635	0.067385445
Column Total	0.102040816	0.013477089	0.115517905

CV	3.841459
P value	0.905727

Contingency table for Larger Sample	Observed		
	High Salt	Normal	Row Total
Died of CVD	23	1977	2000
Died of other	15	1885	1900
Column Total	38	3862	3900

Expected	High Salt	Normal	Row Total
Died of CVD	19.48717949	1980.512821	2000
Died of other	18.51282051	1881.487179	1900
Column Total	38	3862	3900

Chi Square w/ Yates	High Salt	Normal	Row Total
Died of CVD		0.465797908	0.004583201	0.470381109
Died of other		0.490313588	0.004824422	0.495138009
Column Total		0.956111496	0.009407622	0.965519118

CV	3.841459
P value	0.726148

(1)

How to Report a Chi-Square Result

	High Salt	Normal	*Marginal Row Totals*
Died of CVD	5 (4.08) [0.21]	30 (30.92) [0.03]	35
Died of other	2 (2.92) [0.29]	23 (22.08) [0.04]	25
Marginal Column Totals	7	53	60 (Grand Total)

The chi-square statistic is 0.5591. The *p*-value is .45462. *Not* significant at $p < .05$.

The chi-square statistic with Yates correction is 0.1155. The *p*-value is .733947. *Not* significant at $p < .05$.

(2)

Easy Fisher Exact Test Calculator

Success! The Fisher exact test statistic and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results			
	High Salt	Normal	*Marginal Row Totals*
Died of CVD	5	30	35
Died of other	2	23	25
Marginal Column Totals	7	53	60 (Grand Total)

The Fisher exact test statistic value is 0.6882. The result is *not* significant at p < .05.

(2)

How to Report a Chi-Square Result

	High Salt	Normal	*Marginal Row Totals*
Died of CVD	23 (19.49) [0.63]	1977 (1980.51) [0.01]	2000
Died of other	15 (18.51) [0.67]	1885 (1881.49) [0.01]	1900
Marginal Column Totals	38	3862	3900 (Grand Total)

The chi-square statistic is 1.3126. The *p*-value is .251928. *Not* significant at *p* < .05.

The chi-square statistic with Yates correction is 0.9655. The *p*-value is .3258. *Not* significant at *p* < .05.

(2)

Easy Fisher Exact Test Calculator

Success! The Fisher exact test statistic and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results			
	High Salt	Normal	*Marginal Row Totals*
Died of CVD	23	1977	2000
Died of other	15	1885	1900
Marginal Column Totals	38	3862	3900 (Grand Total)

The Fisher exact test statistic value is 0.2595. The result is *not* significant at p < .05.

Start Again

(2)

Project_2_Workings.m

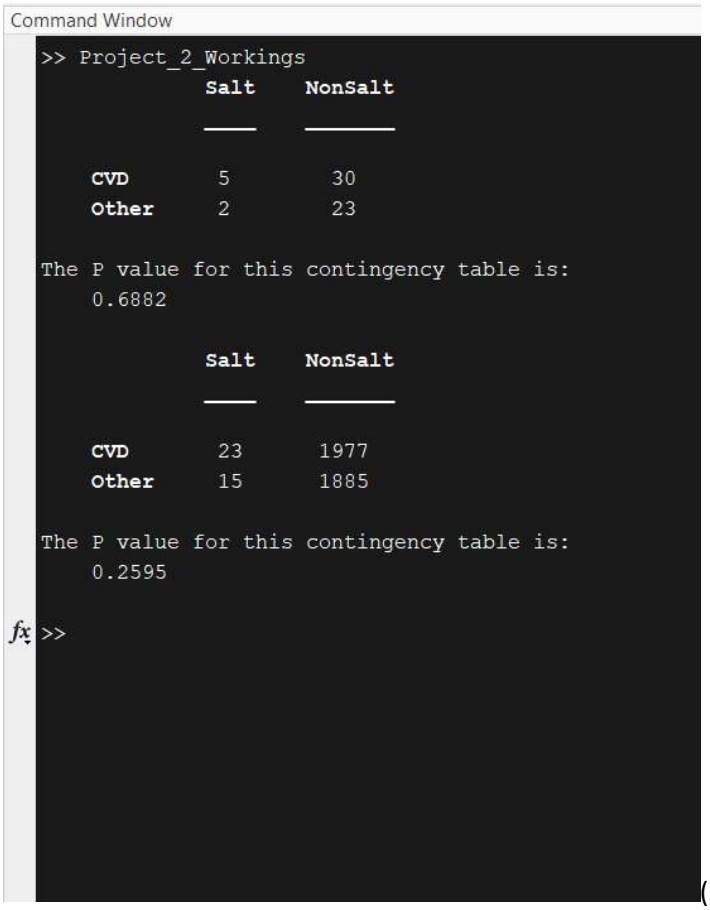
```
x = table ([5;2],[30;23],'VariableNames',{'Salt','NonSalt'},'RowNames',{'CVD','Other'});
[h,p] = fishertest(x);
disp(x);
disp('The P value for this contingency table is:');
disp (p);



y = table ([23;15],[1977;1885],'VariableNames',{'Salt','NonSalt'},'RowNames',{'CVD','Other'});
[H,P] = fishertest(y);
disp(y);
disp('The P value for this contingency table is:');
disp(P);
```

Command Window

(3)



(3)

This table was used to find the critical range in the Tukey Krammer post hoc: https://www.real-statistics.com/statistics-tables/studentized-range-q-table/

This table was used to find the critical range in the Dunnett post hoc: https://www.real-statistics.com/statistics-tables/dunnetts-table/