Biostatistics - Dr. Patrick
BMEN 350; Section 201
Saurabh Dhole 626 002 135
November 19th, 2021

Project 5

A. Read and Reflect

Exclusion criteria are defined as "features of potential study participants who meet inclusion criteria, but possess additional characteristics that may interfere with the success of the study or increase their risk of an unfavorable outcome" [1]. Exclusion criteria are indeed very important when conducting research and clinical studies. Many scientists and clinical researchers would agree that a high validity and high reliability of exclusion criteria can help in curtailing confounding variables and reducing selection bias [2]. Additionally, if there is indeed an association between the exposures or interventions and the outcomes, a high reliability and high validity of exclusion criteria can improve the likelihood of finding said association [2]. In Dr. Corella's informative publication, I understood that the goal of the study was to discern the rate of infection and probability of mortality in non-infected patients who had recently undergone orthopedic and trauma surgeries (OTS) throughout a period of COVID-19 transmission. Dr. Corella and his colleagues state that their exclusion criteria included the presence of symptoms compatible with COVID-19 and the absence of consent to participate in the study. In my opinion, this exclusion criteria makes sense because if Dr. Corella and his colleagues wanted to determine the infection rate and mortality probability in non-infected patients who have recently undergone OTS, introducing patients who are already infected or possibly infected with COVID-19 would indeed be confounding and counterintuitive to the goals of the study. As a Biomedical Engineer, it is indeed very important for me to understand and respect the importance of exclusion criteria. I say this because the application of proper exclusion criteria is crucial to my current undergraduate research project in lymphatic cancer metastasis. I must ensure that I exclude anti-cancer drugs that are irrelevant to the metastasis of certain head and neck cancers. Doing so will allow the laboratory to gauge the effectiveness of the anti-cancer drugs that do indeed curtail cancer cell metastasis in the lymphatic system. In fact, this type of exclusion criteria for testing anti-cancer drugs in the lymphatic system was proposed to my team by the Principal Investigator of the laboratory. Thus, it is followed carefully. In a more general sense, it is indeed important for me to understand and respect the importance of exclusion criteria. This is because I plan on becoming a Physician and conducting biomedical device implant research. During clinical trials, I will certainly need to exclude patients who may not even benefit from the implant. I will also need to exclude patients who do not consent to the implantation of the biomedical device. I plan on keeping this information about the importance of exclusion criteria on my statistical tool belt, so that I can wield it during internships in quality engineering and research projects. I will indeed be able to contribute to discussions with my Principal Investigator on which anti-cancer drugs we should include in the studies.

The Kaplan-Meier survival curves and their associated statistics serve to fulfill the goal of estimating the survival function. A Kaplan-Meier survival curve usually has units of time on the x-axis and percentage survival on the y-axis. In addition, a Kaplan-Meier curve plot will appear as a series of step changes, where the steps occur at times of known participant deaths. In the informative publication by Dr. Klose, it appears to me that the goal of the study was to assess the

oncological outcomes (long term) of patients with left-sided vs right-sided colon cancer of stages I-III. Dr. Klose and his colleagues concluded that there was "no significant negative impact on overall, disease-specific, or relative survival" in patients with left-sided vs right-sided colon cancer. I observed that this conclusion made by Dr. Klose and his colleagues certainly does reflect the Kaplan-Meier survival curves that are presented in the esteemed publication. I could clearly see that the survival functions for the left-sided and right-sided colon cancer patients resembled one another in the overall, disease-specific, and relative survival models. As a Biomedical Engineer, the use of Kaplan-Meier survival curves and their associated statistics is very important to me. I say this because if I am tasked with comparing the effectiveness of a treatment or intervention during a pharmacology internship, I will definitely need to know how to interpret Kaplan-Meier survival curves and their associated statistics. For example, I may need to compile data on the clinical trials of drug A vs drug B, and decide which drug provides the more favorable survival curve for the diseased patients who used the drugs. I can definitely foresee myself using Kaplan-Meier curves and their associated statistics in such pharmacology or quality engineering internships. In a more general sense, the use of Kaplan-Meier curves and their associated statistics are essential for me. I say this because I plan on developing biomedical implant coatings as a Physician. I will definitely need to make sure I develop an implant coating that has a greater median survival time than the median survival time of implant coatings currently on the market. I plan on applying this information on how to use Kaplan-Meier survival curves and their associated statistics to my BMEN 350 module projects as well as quality engineering internships.

The ability to read and understand clinical studies is certainly something that is fundamental for many healthcare professionals and biostatisticians. Physicians and biostatisticians alike may have to research and understand prior clinical studies so as to be able to implement relevant methods and techniques during their own clinical studies. When I first joined the lab headed by my Principal Investigator, I was encouraged to read a few relevant scholarly peer reviewed articles regarding the research area I would be working in. Upon reading the eloquently written publication by Dr. Ranganathan, I learned some key information that will help me read and understand clinical studies better. The key information that I learned from Dr. Ranganathan's publication was the fact that statistical significance does not necessarily imply clinical significance. Dr. Ranganathan cites that clinical significance pertains to implications of a result on existing practice or treatment. I observed in this article that Dr. Ranganathan stated "while there are established, traditionally, accepted values for statistical significance testing, this is lacking for evaluating clinical significance". Over the past twelve weeks in BMEN 350 lecture, I have learned that statistical significance is heavily dependent on factors such as a study's sample size. However, it is up to the interpretation of the Principal Investigator or reader regarding how such factors relate to clinical significance and implications on current treatments. This information greatly affects my ability to read and understand clinical studies as I will now avoid jumping to conclusions in regards to interpreting statistical significance as it relates to clinical significance. My newly bolstered ability to read and understand clinical studies is very important to me as a Biomedical Engineer. I say this because if I am interning for a hospital's statistics department, I will be wary of arriving at clinical conclusions based on the conclusions of statistical significance tests. This practice of caution may indeed be seen favorably by my supervisors. In a general sense, my newly bolstered ability to read and understand clinical studies is indeed crucial. I say this because I am tasked with reading clinical studies frequently during my undergraduate research in lymphatic cancer metastasis. Often times I will be reading clinical studies on anti-cancer drugs and immune-enhancing agents.

Sometimes when reading these studies, I get a little too optimistic about the usage of a certain anti-cancer drug based on its results in statistical significance tests. From now on, I will remind myself to not associate the results of the statistical significance of the drugs to their supposed clinical significance, as the clinical significance is reliant on more than just sample sizes and confidence intervals. I plan on retaining my bolstered ability to read and understand clinical studies for the rest of my life! I say this because it is certainly useful for Engineers and Physicians alike to be able to grasp the details and concepts of pivotal clinical studies. I plan on becoming a Physician, so the safety and well-being of the patients will certainly depend on my ability to read, understand, and critically think about clinical studies that are relevant to the patients.

B. Problem 1

We are tasked with determining if Captopril is able to decrease the presence of urinary protein. To execute this, a paired t-test was performed. A paired t-test was selected for this task because the urinary protein (g / 24hr) levels, of the diabetic nephropathy patients, were measured before and after the 8-week Captopril (37.5 mg) intervention period. The measurements in the provided Captopril data set are indeed paired measurements. Each subject had been observed before and after the treatment. This implies that we are investigating whether or not there is a difference in a single group between two points in time. This is why a paired t-test was applied. More specifically, this t-test was parametric. This is because the measured differences between the before and after data are indeed normally distributed, as the mean and median of the measured differences between the before and after data were very close to one another (mean = 5.84 g / 24hr and median = 5.05 g / 24hr). There were also no outliers among the before or after data. This is because all the measurements were determined to be well within the upper and lower bounds of the before and after data. Since the measured differences between the before and after data are normally distributed, and since the data values are paired measurements (i.e., before and after), the paired t-test was the method of choice to determine whether or not Captopril is able to decrease the presence of urinary protein (please refer to Appendix for calculations).

The null hypothesis (H_o) was established as follows: “there is no difference between the means of the urinary protein levels of the before and after data for the diabetic nephropathy patients”. The alternative hypothesis (H_1) was established as follows: “there is a difference between the means of the urinary protein levels of the before and after data for the diabetic nephropathy patients”. The paired t-test was performed and the p-value that was received was 0.0068. This p-value is indeed less than the alpha value of 0.05. Since p-value is sufficiently less than the alpha, we must reject the null hypothesis (H_o) that there is no difference between the means of the urinary protein levels of the before and after data for the diabetic nephropathy patients.

What can be implied from the rejection of the above null hypothesis? The rejection of the above null hypothesis implies that we do indeed have reason to believe that there is a detectable difference between the means of the urinary protein levels of the before and after data for the diabetic nephropathy patients. From this rejection of the rejection of the null hypothesis, we can conjecture that the decrease in mean urinary protein levels from the before to after data may indeed have been brought about by the presence of the Captopril drug during the 8-week period, given of course that there were no other confounding variables at play throughout the 8-week Captopril intervention period.

C. *Problem 2*

A matrix was established using the data given in the problem. This matrix can be seen in the appendix. A McNemar test was performed for this problem for the following reasons. The data given was paired and nominal. In the given data, there was one nominal variable with two categories (AKA dichotomous variables) and there was one independent variable with connected groups. The McNemar test is non-parametric. The McNemar test was performed with the matrix provided in the appendix, and a p-value of 0.0291 was obtained. This p-value is indeed less than the alpha of 0.05.

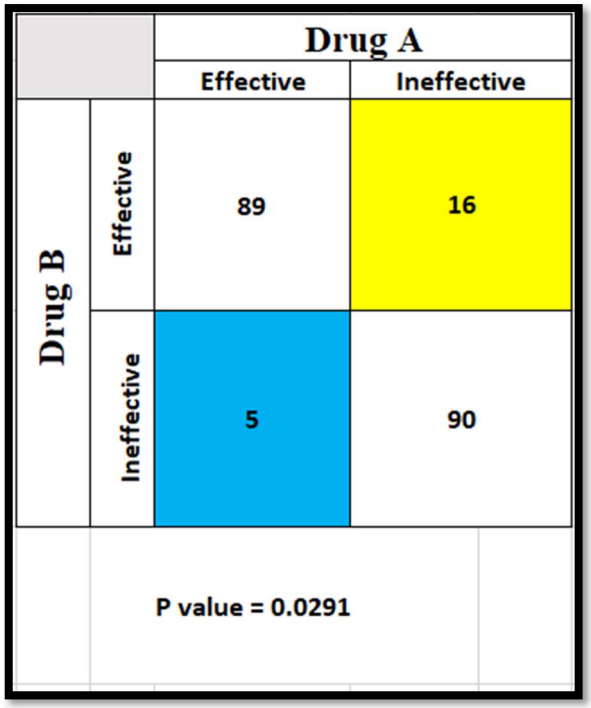


Figure 1: Drug Matrix: This is the matrix that was created using the data provided in problem 2. A McNemar test was performed and the P value that was obtained was 0.0291. The screenshot of the data from the software is provided in the appendix. This P value was greater than the alpha value of 0.05.

The hypothesis states that Drug A is no better than Drug B. We are asked to determine whether or not the data support this hypothesis. The phrase "no better than", according to oxford languages (google search), means "**just (or almost) the same as** (something bad)". Therefore, this hypothesis is proposing that there is no difference between Drug A and Drug B. However, the p-value is less than the alpha. Since the p-value is less than the alpha, we have reason to believe that there is indeed a significant difference between the yellow-colored cell and the blue-colored cell. In other words, there are significantly more individuals (16) in the pairs who stated that Drug A was ineffective and Drug B was effective than there are individuals (5) in the pairs who stated that Drug B was ineffective and Drug A was effective. Therefore, the hypothesis that states that "Drug A is no better than Drug B" isn't exactly correct.

D. Problem 3

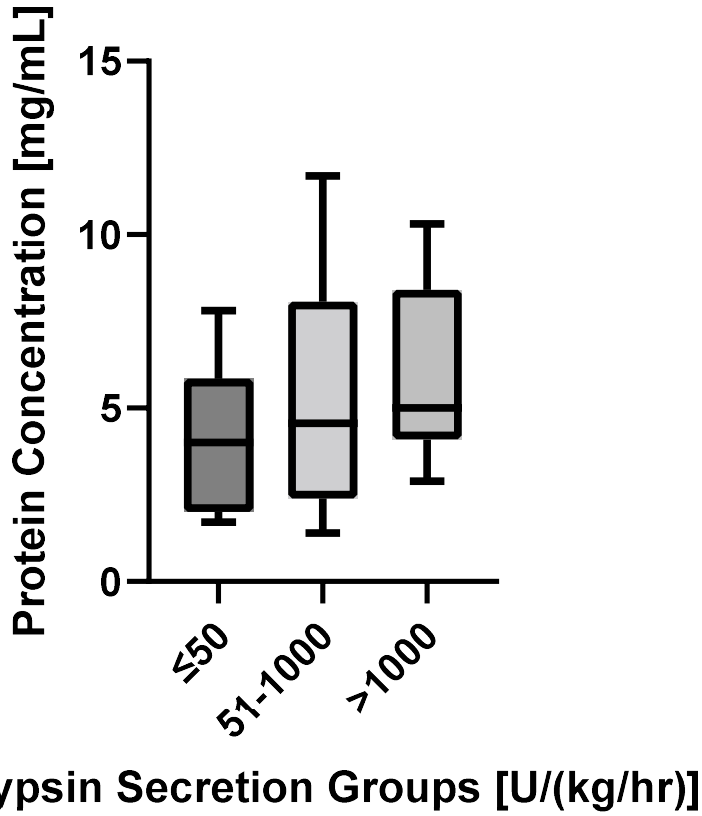


Figure 2: Trypsin Output (low, intermediate, high) vs Protein Concentration: This figure displays box plots of the three Trypsin secretion/output groups. The y-axis displays the protein concentration in mg/mL of each Trypsin secretion/output group. The low Trypsin secretion group (≤ 50 ($U/(\frac{kg}{hr})$) has a mean protein concentration of 3.93 mg / mL $\pm$ 2.21 mg / mL and a median protein concentration of 4.00 mg / mL. The intermediate Trypsin secretion group (51-1000 ($U/(\frac{kg}{hr})$) has a mean protein concentration of 5.41 mg / mL $\pm$ 3.38 mg / mL and a median protein concentration of 4.55 mg / mL. The high Trypsin secretion group (> 1000 ($U/(\frac{kg}{hr})$) has a mean protein concentration of 5.94 mg / mL $\pm$ 2.54 mg / mL and a median protein concentration of 5.00 mg / mL.

Just based off of the box plots, I cannot perceive any major differences between the three groups aside from the fact that the intermediate trypsin secretion group appears to have the highest IQR out of the three. It appears that the low Trypsin secretion group (≤ 50 ($U/(\frac{kg}{hr})$) has the lowest median protein concentration out of the group of three. It also appears that the intermediate Trypsin secretion group (51-1000 ($U/(\frac{kg}{hr})$) has a higher median protein concentration than that of the low Trypsin secretion group (≤ 50 ($U/(\frac{kg}{hr})$), but has a lower median protein concentration than the high Trypsin secretion group (> 1000 ($U/(\frac{kg}{hr})$). The high Trypsin secretion group (> 1000 ($U/(\frac{kg}{hr})$) appears to have the highest median protein concentration out of the three groups. One may indeed mistake these three box plots to be normal distributions as the means and medians of the individual Trypsin secretion groups are relatively close together. The medians of these three groups are indeed different, but are all hovering around the 5 mg/mL mark for protein concentration. The three groups look similar in terms of mean and median. Therefore, I cannot perceive any major differences between these three groups.

Since the question stem states that we should not assume normality, a non-parametric test needs to be used. This non-parametric test is called the Kruskal Wallis test. It is similar to the ANOVA which is usually used for parametric data sets. The null hypothesis (H_o) was established as follows: "There are no differences between or among the median protein concentration of the three Trypsin Secretion groups". The alternative hypothesis (H_1) was established as follows: "There is at least one difference between or among the median protein concentration of these three Trypsin Secretion groups". The Kruskal Wallis non-parametric test was performed and a p-value of 0.2391 was received. This p-value is indeed greater than the alpha value of 0.05. Therefore, we must fail to reject the null hypothesis. We can conjecture that we do not have reason to believe that there is at least one difference between or among the median protein concentration of these three Trypsin Secretion groups. Since the p-value was greater than alpha, there was no need to determine exactly which Trypsin Secretion group had a higher or lower median protein concentration than another Trypsin Secretion group, but never the less the Dunn's post hoc test results are provided in the Appendix.

E. Problem 4

It can be seen in Figure 3 caption below, that the median survival time for Drug 1, Drug 2, and No Drug treatments was determined to be 23 months, 18 months, and 5 months respectively. There was a total of 10 deaths among the patients who were on Drug 1. There was a total of 14 deaths among the patients who were on Drug 2. There was a total of 17 deaths among the patients who were in the No Drug group.

The three survival curves were compared. The null hypothesis was established as follows: "The three different drug treatments have the same effect on survival". The alternative hypothesis was established as follows: "The three different drug treatments have different effects on survival". The log-rank test is the gold standard for comparing survival functions. It equally weighs deaths at all time points. The Gehan-Breslow-Wilcoxon test is also sometimes used for comparing survival curves, but it gives more weight to deaths at earlier time points. The p value obtained from the log-rank test was less than 0.0001 which is indeed less than the alpha of 0.05. The p value obtained from the Gehan-Breslow-Wilcoxon test was also less than 0.0001 which is indeed less than the alpha of 0.05. Since the p values obtained from the log-rank and the Gehan-Breslow-Wilcoxon tests were both less than the alpha of 0.05, we must reject the null hypothesis that "the three different drug treatments have the same effect on survival". We have reason to believe that the three different drug treatments do not have the same effect on survival. The curves are indeed statistically significant from one another (Please see appendix for calculations of P values).

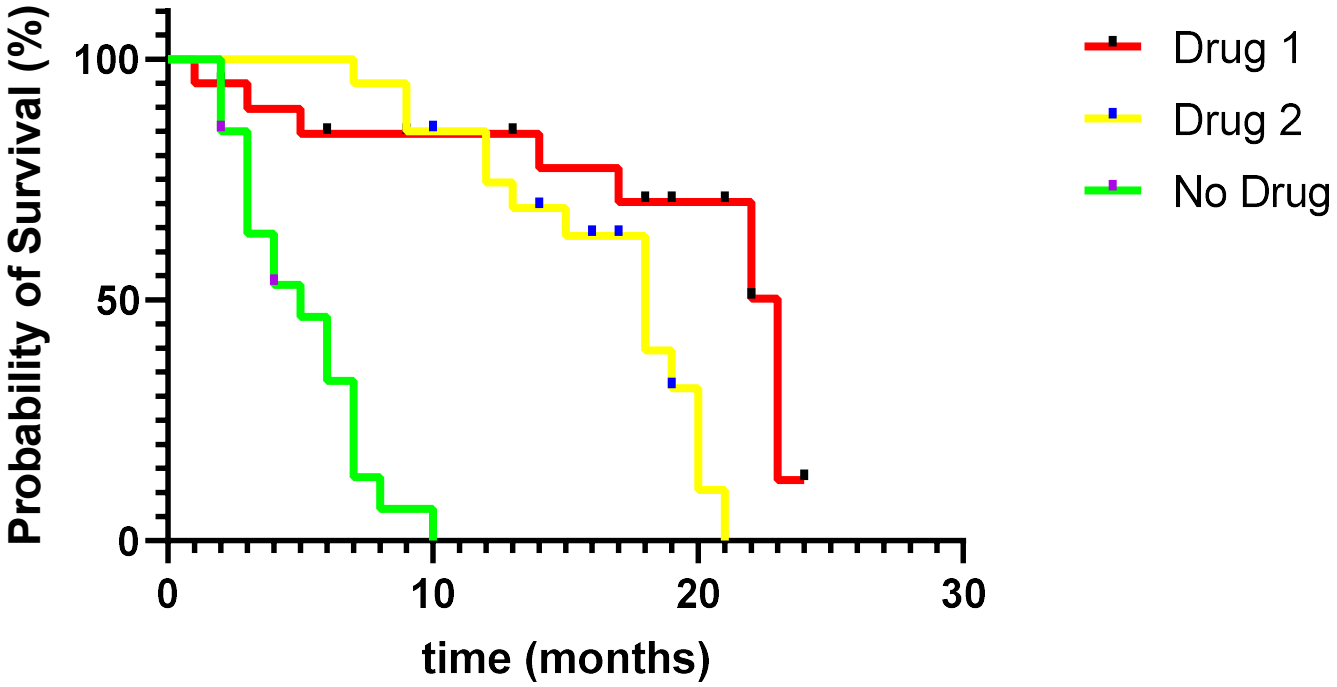


Figure 3: Pancreatic Cancer Kaplan-Meier Survival Curves for Drug 1, Drug 2, and No Drug: The survival curve or survival function for Drug 1 is displayed above as a red line. The black notches on this red line indicate points of censoring (total of 10 censoring events). The median survival time of Drug 1 was determined to be 23 months. There was a total of 10 deaths among the pancreatic cancer patients who were on Drug 1. The survival curve or survival function for Drug 2 is displayed above as a yellow line. The blue notches on this yellow line indicate points of censoring (total of 6 censoring events). The median survival time of Drug 2 was determined to be 18 months. There was a total of 14 deaths among the pancreatic cancer patients who were on Drug 2. The survival curve or survival function for the No Drug group is displayed as a green line. The purple notches on this green line indicate points of censoring (total of 3 censoring events). There was a total of 17 deaths among the pancreatic cancer patients who were in the No Drug list. The median survival time of the No Drug group was determined to be 5 months. Please see appendix for calculations.

The caption for Figure 3 clearly explains that the Drug 1 group saw a total of 10 censoring events. The Drug 2 group saw a total of 6 censoring events, and the No Drug group saw a total of 3 censoring events. A censoring event occurs when participants drop out of the study, the study reaches a pre-determined end time and some participants have survived until the end, or when a certain number of the participants have died. In this case, it is most likely that the patients diagnosed with pancreatic cancer dropped out of the study. However, it can be seen at the end of the red line in Figure 3 (Drug 1) that there is a censoring event at the very end of the red line. This censoring event most likely implies that the study had reached a pre-determined end time and that some of the pancreatic cancer patients had indeed survived!

The survival rates for 6, 12, and 24 months for the Drug 1 group are as follows: The 6-month survival rate for Drug 1 group is 85%. The 12-month survival rate for Drug 1 group is 85%. The 24-month survival rate for Drug 1 group is 15%.

The survival rates for 6, 12, and 24 months for the Drug 2 group are as follows: The 6-month survival rate for Drug 2 group is 100%. The 12-month survival rate for Drug 2 group is 85%. The 24-month survival rate for Drug 2 group is 0%.

The survival rates for 6, 12, and 24 months for the No Drug group are as follows: The 6-month survival rate for No Drug group is 35%. The 12-month survival rate for No Drug group is 0%. The 24-month survival rate for No Drug group is 0%.

References for Reflections portion:

[1] Patino, C. M., & Ferreira, J. C. (2018). Inclusion and exclusion criteria in research studies: definitions and why they matter. *Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, *44*(2), 84. https://doi.org/10.1590/s1806-37562018000000088

[2] Salkind, N. J. (2010). *Encyclopedia of research design* (Vols. 1-0). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412961288

Appendix begins on the next page …

F. Appendix:

Problem 1:

Before	
Mean	10.61
Standard Error	2.168842497
Median	8.05
Mode	#N/A
Standard Deviation	6.858482177
Sample Variance	47.03877778
Kurtosis	1.234120296
Skewness	1.345533208
Range	20.9
Minimum	4.7
Maximum	25.6
Sum	106.1
Count	10
Confidence Level(95.0%)	4.90626259

Q1	5.5
Q3	14.6
IQR	9.1
L Bound	-8.15
U Bound	28.25

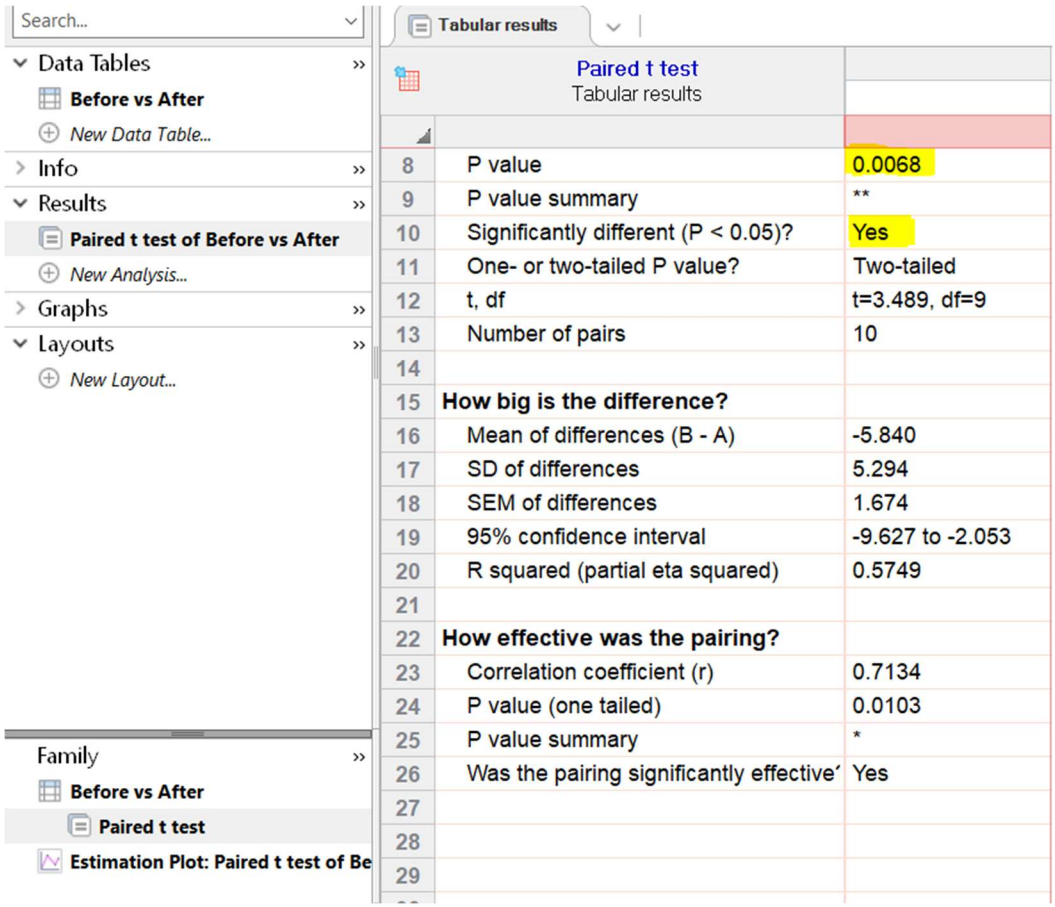
After	
Mean	4.77
Standard Error	0.845517593
Median	5.15
Mode	#N/A
Standard Deviation	2.673761395
Sample Variance	7.149
Kurtosis	0.654533534
Skewness	0.454434239
Range	9.4
Minimum	0.7
Maximum	10.1
Sum	47.7
Count	10
Confidence Level(95.0%)	1.912693679

Q1	3.025
Q3	6
IQR	2.975
L Bound	-1.4375
U Bound	10.4625

Before - After set
15.5
11.3
10.4
7
1.7
7.2
-0.3
0.7
3.1
1.8

Before - After set	
Mean	5.84
Standard Error	1.673997212
Median	5.05
Mode	#N/A
Standard Deviation	5.293643988
Sample Variance	28.02266667
Kurtosis	-0.79007871
Skewness	0.597675557
Range	15.8
Minimum	-0.3
Maximum	15.5
Sum	58.4
Count	10
Confidence Level(95.0%)	3.786844784

Q1	1.725
Q3	9.6
IQR	7.875
L Bound	-10.0875
U Bound	21.4125



Had to check first to see if the measured differences were normally distributed. The measured differences were indeed normally distributed as seen here.

Then I proceeded with the paired t-test.

Problem 2:

Results of McNemar's test for a case-control study

Summary:

If there were no association between the risk factor and the disease, you'd expect the number of pairs where cases was exposed to the risk factor but control was not to equal the number of pairs where the control was exposed to the risk factor but the case did not. In this study, there were 21 discordant pairs (case and control had different exposure to the risk factor). There were 5 (23.810%) pairs where the control was exposed to the risk factor but the case was not, and 16 (76.190%) pairs where the case was exposed to the risk factor but the control was not.

P Value:

The two-tailed P value equals 0.0291

By conventional criteria, this difference is considered to be statistically significant.

The P value was calculated with McNemar's test with the continuity correction.
Chi squared equals 4.762 with 1 degrees of freedom.

The P value answers this question: If there is no association between risk factor and disease, what is the probability of observing such a large discrepancy (or larger) between the number of the two kinds of discordant pairs? A small P value is evidence that there is an association between risk factor and disease.

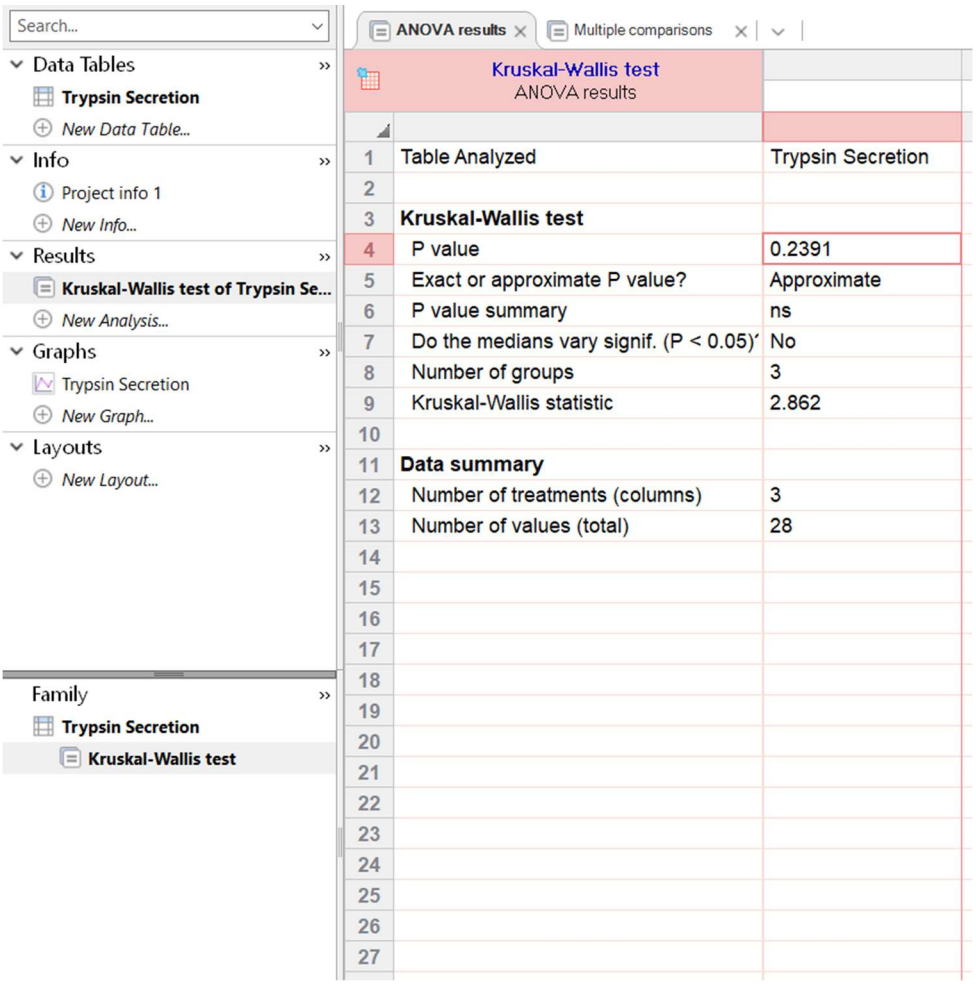
Odds ratio:

The odds ratio is 3.200, with a 95% confidence interval extending from 1.120 to 11.169

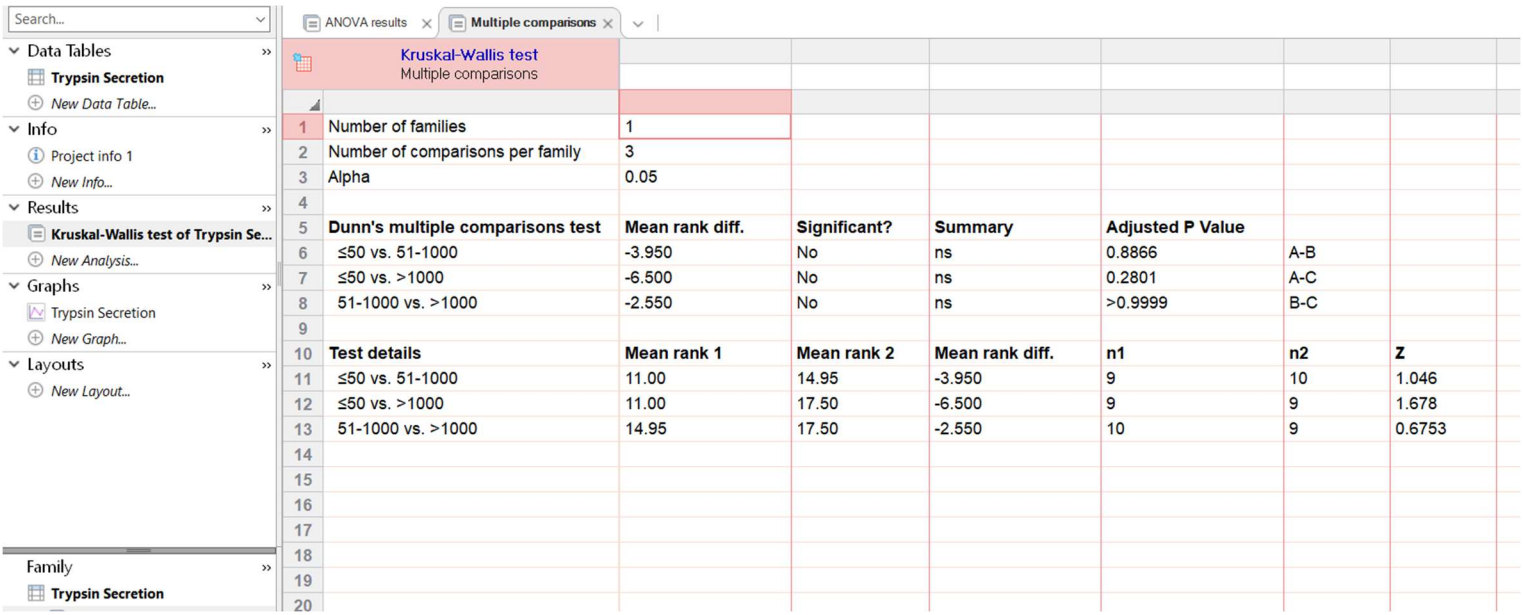
Review your data:

		Control		
		+	-	**Total**
Case	+	89	16	**105**
	-	5	90	**95**
	Total	**94**	**106**	**200**

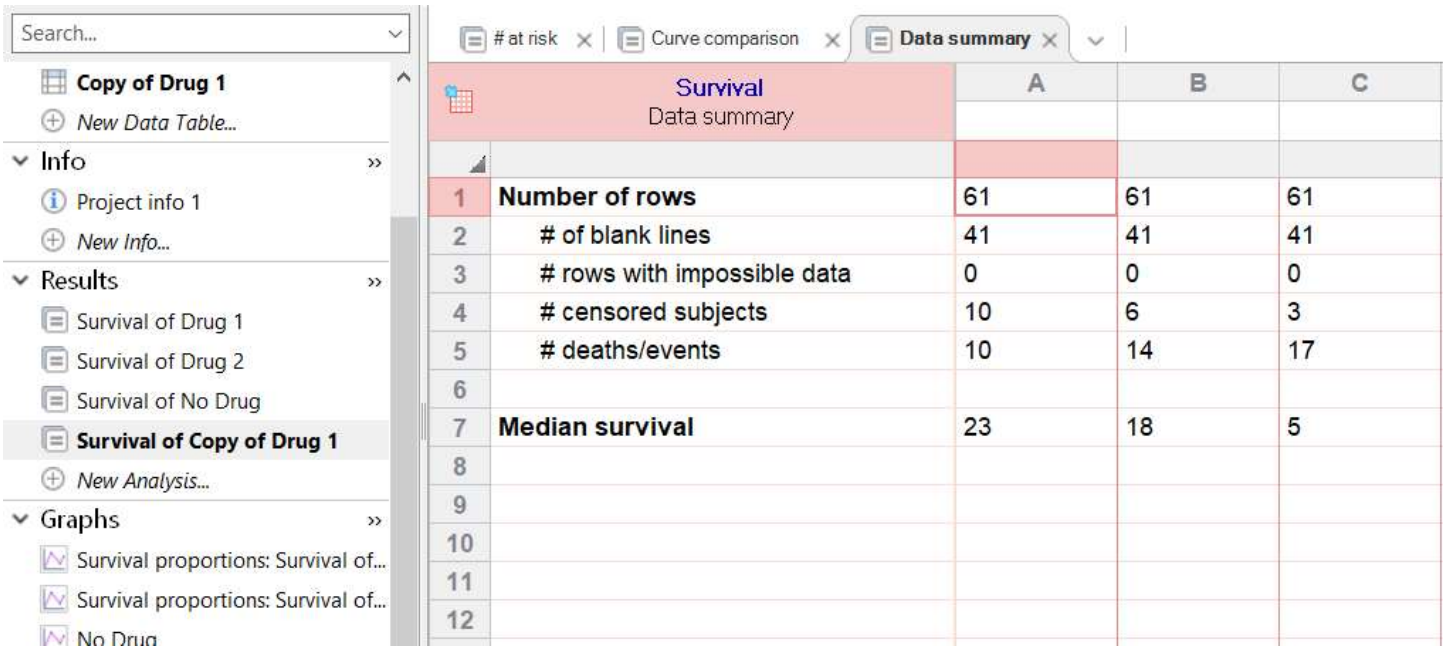
Problem 3:

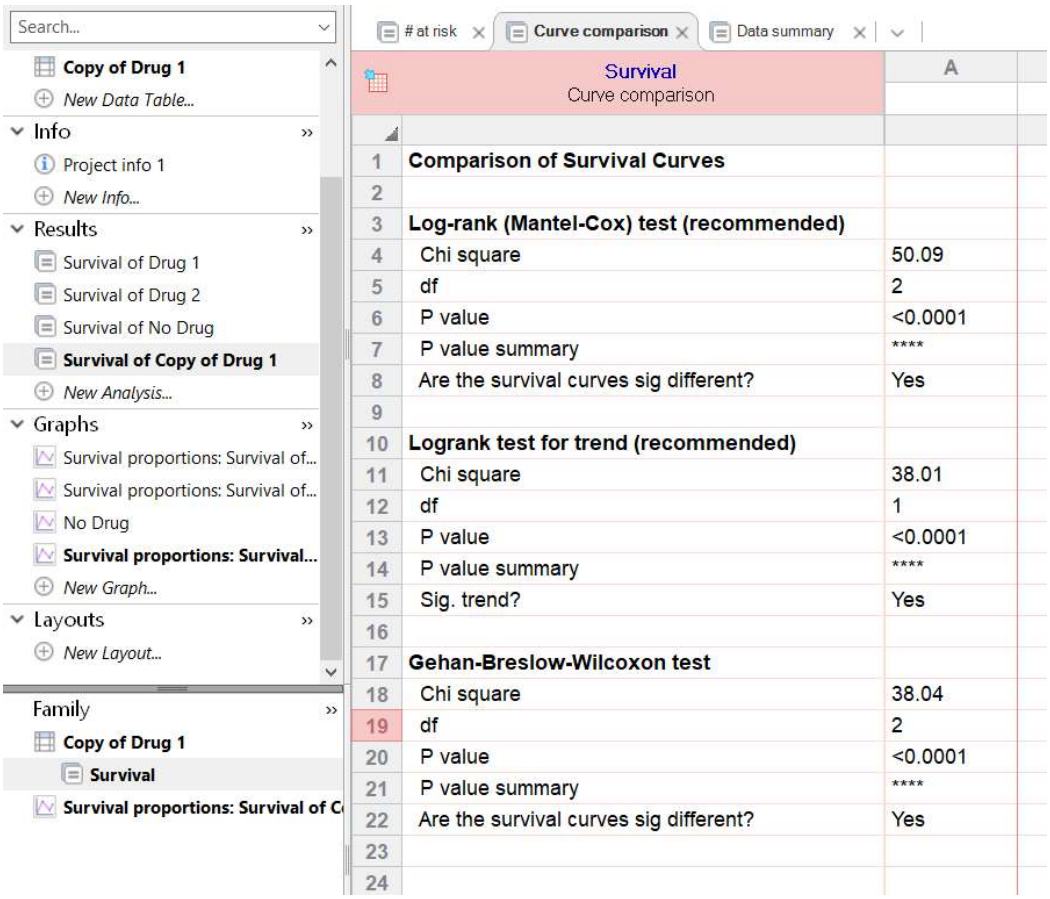


The p value was less than alpha. So, the null hypothesis had to be rejected. But the Dunn's test was performed anyways.



Problem 4:


Search...
Copy of Drug 1
New Data Table...
Info
Project info 1
New Info...
Results
Survival of Drug 1
Survival of Drug 2
Survival of No Drug
Survival of Copy of Drug 1
New Analysis...
Graphs
Survival proportions: Survival of...
Survival proportions: Survival of...
No Drug
at risk
Curve comparison
Data summary
Survival
Data summary
A
B
C
1 Number of rows 61 61 61
2 # of blank lines 41 41 41
3 # rows with impossible data 0 0 0
4 # censored subjects 10 6 3
5 # deaths/events 10 14 17
7 Median survival 23 18 5


Search...
Copy of Drug 1
New Data Table...
Info
Project info 1
New Info...
Results
Survival of Drug 1
Survival of Drug 2
Survival of No Drug
Survival of Copy of Drug 1
New Analysis...
Graphs
Survival proportions: Survival of...
Survival proportions: Survival of...
No Drug
Survival proportions: Survival...
New Graph...
Layouts
New Layout...
Family
Copy of Drug 1
Survival
Survival proportions: Survival of C
at risk
Curve comparison
Data summary
Survival
Curve comparison
A
1 Comparison of Survival Curves
3 Log-rank (Mantel-Cox) test (recommended)
4 Chi square 50.09
5 df 2
6 P value <0.0001
7 P value summary ****
8 Are the survival curves sig different? Yes
10 Logrank test for trend (recommended)
11 Chi square 38.01
12 df 1
13 P value <0.0001
14 P value summary ****
15 Sig. trend? Yes
17 Gehan-Breslow-Wilcoxon test
18 Chi square 38.04
19 df 2
20 P value <0.0001
21 P value summary ****
22 Are the survival curves sig different? Yes