BME 423 Fall 2021: Homework 4

Due by 5 PM on Friday, December 3rd, 2021. Please read the homework guidelines before starting your assignment. Statistical tables necessary for completing this assignment are posted on Blackboard.

1. [20 points]

A physician was interested in the effect of an anesthetic on reaction times. Two groups were compared: one group of 14 subjects was given the anesthetic, while the other group of 14 subjects was given a placebo. Subjects had to press a button when they identified a specific visual stimulus and their reaction time (the time between when the stimulus was shown and when the subject pressed the button) was measured in milliseconds. The table below contains the data for the two groups.

With Anesthetic	Without Anesthetic
135	129
141	131
145	138
165	139
167	142
173	143
178	144
191	147
244	155
245	156
256	163
267	171
268	192
282	230

Do these data support the hypothesis that there is a difference in reaction time between people given the anesthetic and people not exposed to it? Use an appropriate nonparametric test and the corresponding parametric test. Discuss any differences in the two conclusions. Note: you still need to "test" the null hypotheses, even though the question was phrased using an alternative hypothesis.

H_0: There is no difference in reaction time between the placebo and anesthetic groups.

$\alpha = 0.05$

NONPARAMETRIC TEST: Mann-Whitney Rank-Sum Test

bigger With Anesthetic	smaller Without Anesthetic
135 3	129 1
141 6	131 2
145 10	138 4
165 15	139 5
167 16	142 7
173 18	143 8
178 19	144 9
191 20	147 11
244 23	155 12
245 24	156 13
256 25	163 14
267 26	171 17
268 27	192 21
282 28	230 22

$$T = \sum_{i=1}^{n_s} r_i = 146$$

$$\mu_T = \frac{n_s(n_s + n_B + 1)}{2} = \frac{14(28+1)}{2} = 203$$

$$\sigma_T = \sqrt{\frac{n_s n_B (n_s + n_B + 1)}{12}} = \sqrt{\frac{14 \cdot 14 (29)}{12}}$$

$$= 21.764$$

$$z_T = \frac{|T - \mu_T| - \frac{1}{2}}{\sigma_T} = \frac{|146 - 203| - \frac{1}{2}}{21.764} = 2.596$$

$z_{crit} = t_{0.05}(\nu \to \infty) = 1.96$ $\qquad$ $z_{0.01}(\nu \to \infty) = 2.576$

$4.434 > 2.576 > 1.96$

$\therefore$ There is a significant difference in reaction time between the placebo and anesthetic groups. ($p < 0.01$)

H_0: There is no difference in reaction time between the placebo and anesthetic groups.

$\alpha = 0.05$

PARAMETRIC TEST: Unpaired t-Test

T = Treatment
C = Control

n_T With Anesthetic	n_c Without Anesthetic
135	129
141	131
145	138
165	139
167	142
173	143
178	144
191	147
244	155
245	156
256	163
267	171
268	192
282	230

$$\bar{x}_T = \frac{1}{n_T}\sum_{i=1}^{n_T} x_i = 204.0714 \qquad \bar{x}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} x_i = 155.7143$$

$$s_T^2 = \frac{1}{n_T-1}\sum_{i=1}^{n_T}(x_i-\bar{x}_T)^2 = 2849.302 \qquad s_c^2 = \frac{1}{n_c-1}\sum_{i=1}^{n_c}(x_i-\bar{x}_c)^2 = 738.681$$

$$s^2 = \frac{(n_c-1)s_c^2 + (n_T-1)s_T^2}{n_c+n_T-2} = \frac{s_c^2+s_T^2}{2} = 1793.992$$

$\nu = n_c + n_T - 2$
$= 26$

$$t = \frac{\bar{x}_T - \bar{x}_c}{s\sqrt{\frac{1}{n_c}+\frac{1}{n_T}}} = \frac{204.0714 - 155.7143}{\sqrt{1793.992\left(\frac{1}{14}+\frac{1}{14}\right)}}$$

$= 3.0206$

$t_{0.05}(26) = 2.056$

$t_{0.01}(26) = 2.779$

$3.0206 > 2.779$

$\therefore$ There is a significant difference in reaction time between the placebo and anesthetic groups. $(P < 0.01)$

2. [18 points]

Eleven children with autism were enrolled in a study to assess the effectiveness of a new drug designed to reduce repetitive behaviors in this population. The amount of time (in minutes) that each child spent engaged in repetitive behavior during a three-hour observation period was measured both before treatment and then again after taking the new medication for one week (see table below). It has been determined that these data are <u>not</u> Normally distributed. Is there any evidence of a difference in duration of repetitive behaviors between the two observation periods

Before Treatment	After 1 Week of Treatment
68	47
36	19
74	67
66	43
85	21
88	90
52	35
89	89
71	54
87	76
46	52

Before Treatment	After 1 Week of Treatment	Difference $x_2 - x_1$	|Rank|	Signed rank of |difference|
68	47	−21	8	−8
36	19	−17	6	−6
74	67	−7	3	−3
66	43	−23	9	−9
85	21	−64	10	−10
88	90	+2	1	+1
52	35	−17	6	−6
89	89	0 --	- -	- -
71	54	−17	6	−6
87	76	−11	4	−4
46	52	+6	2	+2

$n = 10$

$$W = \sum_{i=0}^{n_r} r_i = -49$$

H_0: There is no difference in amount of time spent engaging in repetitive behavior between before treatment and after 1 week of treatment.

$\alpha = 0.05$

Test: Wilcoxon Signed-Rank

$$\sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{6} - \sum_{i=1}^{1} \frac{(T_i-1)T_i(T_i+1)}{12}} \qquad T_i = 3$$

$$\frac{(T_i-1)T_i(T_i+1)}{12} = \frac{2(3)(4)}{12} = \frac{24}{12} = 2$$

$$\sigma_w = \sqrt{\frac{10(11)(21)}{6} - 2} = 19.570$$

$$z_w = \frac{|W| - \frac{1}{2}}{\sigma_W} = \frac{49 - \frac{1}{2}}{19.570} = 2.478$$

$$z_{0.05} = 1.96 \qquad z_{0.02} \doteq 2.32$$

$$2.478 > 2.32$$

$\therefore$ There is a significant difference in amount of time spent engaging in repetitive behavior between before treatment and after 1 week of treatment. $(P < 0.02)$

3. [15 points]
A known human parasite that lives primarily in the blood has distinct environmental needs in order to spread from person to person. To test if different regions in the U.S. had different suitabilities for the proliferation and dissemination of this parasite, levels of an antibody specific to an epitope expressed by the parasite were measured in subjects from 3 different geographical regions.

One subject from Region A and two subjects from Region C did not show up for their interview/testing session, dropping the total number of subjects enrolled in the study from 24 down to 21. The collected data are summarized in the table below. Do these data support the hypothesis that there is a difference in the prevalence/infection levels of this parasite across these three regions? You <u>cannot</u> assume that antibody levels are Normally distributed.

Note: you still need to "test" the null hypothesis, even though the question was phrased using an alternative hypothesis.

$n_A = 7$ $n_B = 8$ $n_C = 6$

2 ties between 2 people

Antibody levels (kU/L)

<u>Region A</u>	<u>Region B</u>	<u>Region C</u>
1.1 1	1.8 4	3.9 8.5
1.2 2	4.9 10.5	4.9 10.5
1.5 3	7.2 14	5.5 12
2.1 5	8.1 15	6.6 13
2.2 6	8.9 17	8.2 16
2.9 7	9.7 19	9.1 18
3.9 8.5	11 20	
	15 21	

$\bar{R}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} R_i = 4.64$ $\bar{R}_B = \frac{1}{n_B}\sum_{i=1}^{n_B} R_i = 15.0625$ $\bar{R}_C = \frac{1}{n_C}\sum_{i=1}^{n_C} R_i = 13$

H_0: There is no difference in antibody level between any of the three regions.

$\alpha = 0.05$

Test: Kruskal-Wallis Test

$N = n_A + n_B + n_C = 21$

$$\bar{R} = \frac{N+1}{2} = \frac{21+1}{2} = 11$$

$$D = \sum_{t=1}^{3} n_t(\bar{R}_t - \bar{R})^2 = 7(4.64-11)^2 + 8(15.0625-11)^2 + 6(13-11)^2$$

$$= 438.924$$

$$H = \frac{D}{N(N+1)/12} = \frac{438.924(12)}{21(22)} = 11.4006$$

$\nu = k-1 = 2$

$\chi^2_{0.05}(2) = 5.991$

$\chi^2_{0.005}(2) = 10.597$

correction factor: $1 - \frac{\Sigma(\tau_i - 1)\tau_i(\tau_i + 1)}{N(N^2-1)}$

$\Sigma(\tau_i - 1)\tau_i(\tau_i+1) = 2[1(2)(3)]$

$= 2[6] = 12$

$1 - \frac{12}{21(21^2-1)} = 0.9987$

$\chi^2_{0.05}(2) = 5.991$

$\chi^2_{0.005}(2) = 10.597$

$H = H/cF = \frac{11.4006}{0.9987} = 11.415$

$11.415 > 10.597$

$\therefore$ There is a significant difference in antibody level between any of the three regions. $(p < 0.005)$

4. [15 points]

A study was done on a treatment for preventing additional heart attacks in subjects who have already experienced one or more heart attacks. The treatment was administered as soon as each patient was diagnosed with a heart attack in the emergency room. The table below shows the time between diagnosis/start of treatment and a subsequent heart attack in these patients; censored (loss to follow-up) observations are included (indicated by a "+" next to their entry

a. Construct a survival table (do the calculations manually – do not use R) and show the following columns: time to event; interval number; # without an event at beginning of each interval; # who experienced an event (i.e., a heart attack) at end of interval; fraction who did not experience an event in interval; cumulative fraction without an event.

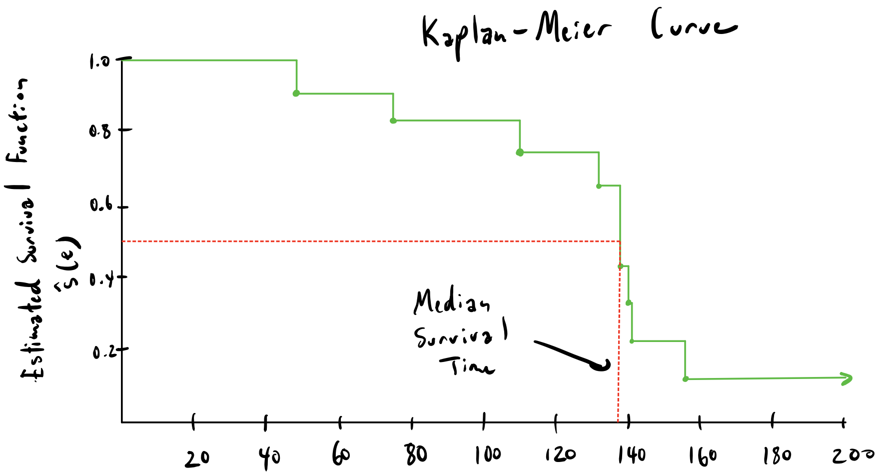
b. Using the data from part (a), draw a Kaplan-Meier plot. What is the median time to event (i.e., what is the median time before another heart attack occurs)?

Patient Number	Time to Event (days)
1	75 2
2	132+
3	110 3
4	137 5.5
5	141 8
6	110+
7	137 5.5
8	185+
9	156 9
10	49 1
11	140 7
12	132 4

$\hat{S}(t_0) = 1$

$\hat{S}(t_i) = F_i \cdot \hat{S}(t_{i-1})$

Patient #	t_i Time to Event	No. alive at beginning of interval n_i	No. of deaths at end of interval (d_i)	$F_i = \frac{n_i - d_i}{n_i}$ Fraction surviving interval	Cumulative Survival Rate $\hat{S}(t)$
10	49	12	1	0.9167	0.9167
1	75	11	1	0.9091	0.833
3	110	10	1	0.9	0.75
6	110+				
12	132	8	1	0.875	0.65625
2	132+				
4 & 7	137	6	2	0.667	0.4375
11	140	4	1	0.75	0.328125
5	141	3	1	0.667	0.21875
9	156	2	1	0.50	0.109375
8	185+				



Median Survival time $(\hat{S}(t_i) \leq 0.05) = 137$ days

5. [20 points]

In treating patients with HIV-1, in addition to monitoring their viral load (number of copies of mRNA from the virus in the patient's plasma), the patient's CD4 cell count is also measured (CD4 cells are a type of lymphocyte that the human immunodeficiency virus kills). A study was conducted to assess the time to return to an acceptable value of CD4 from the start of therapy. Two drug therapy regimens were examined. One group of 17 patients received a two-drug combination of AZT + zalcitabine, while a second group of 17 patients received AZT + zalcitabine + saquinavir (both AZT and zalcitabine are nucleoside analog reverse transcriptase inhibitors; saquinavir is a protease inhibitor).

The data from this study are included in the R package "KMsurv" that was presented in the Discussion Session (data file "drughiv"). In R Studio, review the description of the data file "drughiv" in package "KMsurv". Reference: Klein, John P., and Melvin L. Moeschberger. Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer, 2003.

i) Using R, perform a complete time-to-event (survival) analysis of the time to "CD4 return" for the patients in the AZT + zalcitabine arm of the trial. Show both the resulting survival analysis table, as well as a Kaplan-Meier plot including the 95% confidence region. What is the median time for CD4 return? At this time, what is the 95% confidence interval for the probability of CD4 return? In your homework solution, show your R code as well as the relevant R output, along with the answers to the questions.

ii) Using R, perform a complete time-to-event (survival) analysis of the time to "CD4 return" for the patients in the AZT + zalcitabine + saquinavir arm of the trial. Show both the resulting survival analysis table, as well as a Kaplan-Meier plot including the 95% confidence region. What is the median time for CD4 return? At this time, what is the 95% confidence interval for the probability of CD4 return? In your homework solution, show your R code as well as the relevant R output, along with the answers to the questions.

iii) Using R, compare your estimated survival functions from the two therapies by testing the null hypothesis that they are not different. In your homework solution, show your R code as well as the relevant R output, along with the results of your hypothesis test.

6. [12 points]
Choosing the appropriate statistical test is a critical first step in any biostatistics problem. Please give the name of the statistical test or procedure that should be used for each of the following experiments. The answer to most of the questions may be found in the table on the inside cover of your textbook. You will need to ask yourself the following questions:

i. Are there:
   - 2 groups?
   - 3 or more groups?
   - 2 random variables that may or may not be associated?

ii. If you have groups, do they consist of
   - Different subjects?
   - The same subjects?

iii. Are the data
   - Interval? (e.g., age, height, blood pressure)
     - If so, are they normally distributed?
   - Nominal? (e.g., died/survived)
   - Ordinal? (e.g., pain rating measured from 1 to 5, satisfaction rating measured from 1 to 10)
   - Time-to-event? (e.g., survival)

a. Fifty-two bipolar patients are enrolled in a clinical trial. Half of them are given olanzapine, and half are given lithium. The investigator reports whether there were any seizures during treatment for patients in both groups. After finding a difference between the two groups using an appropriate statistical test, you want to know how much of an association there is between seizures and lithium treatment as compared to olanzapine.
   - Prospective study design
   - Nominal data (two nominal-data outcomes: yes/any or no/none)
   - Looking for association

Relative Risk OR Odds Ratio

b. Eighteen patients with obstructive sleep apnea are fitted with an oral appliance to prevent apnea. Each patient's apnea and hypopnea index (AHI) is measured with and without the appliance. The results are not Normally distributed. You want to know whether there is a difference in AHI with and without the oral appliance.
   - Non-Normal interval data
   - 2 observations in same subjects

Wilcoxon Signed-Rank

c. You are studying a group of cancer patients, and you want to know whether a new medication prolongs their survival time as compared to subjects who receive the standard treatment.
   - Compare time-to-event data

Log-Rank Test

d. Two groups of 13 teenagers each are asked to rank how happy they are after they watch TV (first group) or after they play a video game (second group). You want to know whether there is a difference in the perceived happiness of teenagers after watching TV and after playing a video game.
- Ordinal data
- 2 groups made up of different individuals   Mann-Whitney rank-sum test

e. Migraine patients are divided into three groups and treated preventively for a week with acupuncture (group 1), aspirin (group 2), and indomethacin (group 3). You want to know whether the choice of treatment affects the proportion of patients who develop at least one migraine during the week of the treatment.
- Nominal data (two nominal-data outcomes: yes/any or no/none)
- 3 groups of different subjects χ^2 analysis of contingency

f. The blood pressure of 8 subjects is measured when they are asleep, when they are meditating, and when they are watching The Simpsons. All the data are found to be Normally distributed. You want to know if there is a difference in blood pressure in the three situations.
- Normally-distributed interval data
- 3 measurements in the same subjects   ANOVA test