Homework Assignment III

The questions in this part are based on Chapters 11. Please read through my notes and relevant sections of the book to answer the questions. Good luck!

***Directions* :** *Identify the correct alternative for the following questions. You can put a cross or a tick mark on the chosen alternative.*

Questions 1-4 are based on the following study :

A study conducted in 2007 analyzed whether people belonging to different ethnic groups differed with respect to their political beliefs. We are only given the following incomplete contingency table.

	Political Party Identification			
Ethnic Group	Democrat	Independent	Republican	**Total**
White	516	123		1249
Black	623		576	
Hispanic		75		460
Total		295	1364	

1. Complete the above table. (*There are 7 blanks to fill up).*

2. What will be the conditional distribution of party identification for Hispanics?

a) (38.33%, 46.28%, 45%) b) (45%, 16.3%, 38.7%) c) (6.9%, 2.49%, 5.92%) d) (45%, 16.3%, 25%).

3. The following incomplete table shows the population conditional percentage values for a few cells. Complete the table in such a way that ethnicity is independent of political affiliation.

	Political Party Identification			
Ethnic Group	Democrat	Independent	Republican	**Total**
White				100%
Black	40%			100%
Hispanic			49%	100%

Extra Credit Question

4. Based on Q.1, what would have been the conditional distribution of party identification for Hispanics if ethnicity was independent of political affiliation? *(Hint : Think about expected counts of the cells).*

a) (44.79%, 9.82%, 40%) b) (45%, 16.3%, 38.7%) c) (6.9%, 2.49%, 5.92%) d) (44.79%, 9.82%, 45.39%).

Questions 5-9 are based on the following study :

The following table refers to a study recently carried out by the Florida Transportation Bureau about the effect of wearing seat belts on the seriousness of injury in automobile accidents.

	Type of Injury	
Seat Belt	Serious	Minor
No	3925	18,674
Yes	2150	20,750

Extra Credit Question

5. Find the numerical value and the degrees of freedom of the Chi square test statistic for testing the hypotheses of independence between degree of injury and seat-belt usage.

a) $\chi^2 = 625.98; df = 2$ b) $\chi^2 = 25.65; df = 2$ c) $\chi^2 = 625.98; df = 1$ c) $\chi^2 = 67.58; df = 4$.

6. Based on the p-value of the above chi-square statistic, what should be your conclusion?

a) Seat-belt usage and seriousness of injury are independent at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.
b) Seat-belt usage and seriousness of injury are dependent at $\alpha = 0.1$ but not at $\alpha = 0.05$ and $\alpha = 0.01$.
c) Seat-belt usage and seriousness of injury are dependent at $\alpha = 0.05$ and $\alpha = 0.1$ but not at $\alpha = 0.01$.
d) Seat-belt usage and seriousness of injury are dependent at H_0 at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.

7. Suppose Mr. X is just learning Statistics and does not know anything about Chi-square test. Worse still, he gets confused and mistakenly assumes that "seat-belt usage" is the response while "injury" is the predictor variable. Let p_1 : population proportion of individuals who does not use the seat-belt and have serious injuries and p_2 : population proportion of individuals who use the seat-belt and have serious injuries. If Mr. X tests $H_0 : p_1 = p_2$ against $H_a : p_1 \neq p_2$, what will be the value of the z statistic that he will get? *(Hint : You can use information from Q.5)*

a) 18.09 b) -5.06 c) -25.02 d) 5.06 e) 25.02 f) -18.09.

8. If "seat-belt usage" and "injury" were truly independent in the population, how many more OR less individuals would have minor injuries if they used seat-belts?

a) About 908 individuals less.
b) About 10,306 individuals less.
c) About 908 individuals more.
d) About 10,306 individuals more.
e) Under independence too, there would have been about the same number of individuals who used seat-belts and had minor injuries.

9. How can you interpret the relative risk of serious injuries against seat-belt usage?

a) Estimated proportion of people who had serious injuries wearing seat-belts is about 1.85 times the estimated proportion of those who had serious injuries not wearing seat-belts.
b) Estimated proportion of people who had serious injuries not wearing seat-belts is about 0.54 times the estimated proportion of those who had serious injuries wearing seat-belts.

c) Estimated proportion of people who had serious injuries not wearing seat-belts is about 1.85 times the estimated proportion of those who had serious injuries wearing seat-belts.
d) All of the above.
e) None of the above.

Questions 10-12 are based on the following study :

A general social survey looked at the association pattern between how happy a person is and how many close friends he/she has. Accordingly, we have the following table

	Happiness			
Close Friends	Not too Happy	Pretty Happy	Very Happy	**Total**
0 - 1	52	105	67	224
2 - 5	60	223	165	448
6 or more	37	197	135	369

10. Which of the following statements are true?

a) Estimated probability of being "very happy" is nearly the same for individuals having 2-5 and 6 or more close friends.
b) Estimated probability of being "pretty happy" is about 0.6 higher for individuals having 6 or more close friends than those having only 0-1 close friends.
c) Estimated probability of being unhappy is about 0.13 higher for individuals having 0-1 close friends than those having 6 or more close friends.
d) Estimated probability of being unhappy is about 0.1 higher for individuals having 0-1 close friends than those having 2-5 close friends.
e) (a), (b) and (c) f) (a), (c) and (d) g) All of the above.

Extra Credit Question

11. Which of the following statements are true?

a) Under independence (between number of close friends and happiness), there would have been about 20 more individuals with 0-1 close friends who are "not too happy".
b) Under independence, there would have been about 12 more individuals with 0-1 close friends who are "very happy".
c) Under independence, there would have been about 16 less individuals with 6 or more close friends who are "not too happy".
d) Under independence, there would have been about 11 more individuals with 6 or more close friends who are "pretty happy".
e) All of the above.

12. Suppose the standard error of the residual for the cell corresponding to 0-1 close friends and being "not too happy" is 4.5. Then

a) The observed number of people in this cell is 4.5 standard errors more than what was expected under independence.
b) The observed number of people in this cell is 0.22 standard errors more than what was expected under independence.
c) The observed number of people in this cell is 4.43 standard errors less than what was expected under independence.
d) The observed number of people in this cell is 4.43 standard errors more than what was expected under independence.

13. Suppose we form a new table from the above one by merging columns 2 and 3 and rows 1 and 2. So, our new table would look like

	Happiness		
Close Friends	Not too Happy	Happy	**Total**
0 - 5	112	560	672
6 or more	37	332	369

Based on the odds ratio of being unhappy for the two levels of number of friends, what will be your conclusion ?

a) The odds of being unhappy if you have 6 or more friends is 1.8 times the odds if you have 0-5 friends.
b) The odds of being unhappy if you have 0-5 friends is 0.55 times the odds if you have 6 or more friends.
c) The odds of being unhappy if you have 6 or more friends is 0.11 times the odds if you have 0-5 friends.
d) The odds of being unhappy if you have 0-5 friends is 1.8 times the odds if you have 6 or more friends.

14. Here also Mr. X gets mixed up on the definitions of response and predictor variables and uses a table whose orientation is just the opposite of the above one (happiness as rows, friendship as columns). Based on the new odds ratio, what will he conclude?

a) The odds of being unhappy if you have 6 or more friends is 1.8 times the odds if you have 0-5 friends.
b) The odds of being unhappy if you have 0-5 friends is 0.55 times the odds if you have 6 or more friends.
c) The odds of being unhappy if you have 6 or more friends is 0.11 times the odds if you have 0-5 friends.
d) The odds of being unhappy if you have 0-5 or more friends is 1.8 times the odds if you have 6 or more friends.

15. When do we use Fisher's exact test?

a) When at least one observed cell count for a 2×2 table is less than 5.
b) When the chi-square test statistic for a 2×2 table is too large.
c) When at least one expected cell count for a 2×2 table is less than 5.
d) When the number of rows and/or columns of a contingency table is greater than 2.

The following questions are based on Chapter 15 : Nonparametric Inference

16. When do we use Nonparametric procedures?

a) When the response variable have a skewed distribution in the population.
b) When the sample sizes in one of the groups being compared is too small.
c) When the sample sizes in the groups are too small to even check for normality of the response.
d) When the random sampling assumption is violated.
e) (a) and (c) f) (a) and (b) g) (a), (b) and (c) h) (a), (c) and (d) i) All of the above.

Questions 17-22 are based on the following study :

A study was designed to test whether the educational achievements of a student has anything to do with whether he/she is extrovert or introvert. Accordingly, a researcher at UF took a survey in which the students of a STA 2023 class were told to identify themselves as either introvert or extrovert. The final exam scores of 3 randomly selected students in each group are given below

Extrovert : 85, 68, 91
Introvert : 93, 87, 72

17. Suppose the researcher wants to perform a Wilcoxon test to compare the scores of the two groups of students. What should be the alternative hypotheses ?

a) The expected value of the sample mean rank of the scores of the extrovert group is higher than that of the introvert group.
b) The expected value of the sample mean rank of the scores of the extrovert group is lower than that of the introvert group.
c) The expected value of the sample mean rank of the scores of the introvert group is different than that of the extrovert group.
d) The expected values of the sample mean rank of the scores of the introvert and extrovert groups are the same.
e) The expected value of the sample median rank of the scores of the extrovert group is lower than that of the introvert group.

18. What will be the value of the test statistic and the corresponding p-value ? *(You can put a rank of 1 to the highest score and so on).*

a) Test statistic = 3; p value = 1/20.
b) Test statistic = 1; p value = 7/20.
c) Test statistic = 3; p value = 1/10.
d) Test statistic = 1; p value = 7/10.
e) Test statistic = - 3; p value = 1/10.
f) Test statistic = - 1; p value = 7/10.

19. Suppose Mr. X starts off by assigning a rank of 1 to the lowest score instead. Then, based on the p-value, what will be his conclusion?

a) Reject H_0 at $\alpha = 0.1$ but not at $\alpha = 0.05$ and $\alpha = 0.01$.
b) Reject H_0 at $\alpha = 0.05$ and $\alpha = 0.1$ but not at $\alpha = 0.01$.
c) Reject H_0 at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.
d) Fail to reject H_0 at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.

20. Suppose instead of the Wilcoxon test, we perform a Wilcoxon rank sum test for the above problem. Our test statistic can be either the sum of the ranks for the extrovert or the introvert groups. What will be the p-value?

a) 1/10 b) 7/10 c) 1/20 d) 7/20 e) 5/10

Extra Credit Question

21. For both the Wilcoxon test and Wilcoxon rank sum test, which of the following combination of ranks would give the lowest achievable p-value ? *(Hint : You can get the lowest achievable p-value from the sampling distribution of the scores).*

a) Extrovert : (3, 1, 5); Introvert : (6, 4, 2).
b) Extrovert : (1, 2, 5); Introvert : (3, 4, 6).
c) Extrovert : (1, 2, 3); Introvert : (4, 5, 6).
d) Extrovert : (6, 5, 1); Introvert : (2, 3, 4).
e) Extrovert : (3, 1, 6); Introvert : (2, 4, 5).

22. Suppose the researcher doubles the sample size and now has 6 students in each of the introvert and extrovert groups. Their scores are as follows

Extrovert : 85, 68, 91, 85, 72, 83
Introvert : 91, 87, 72, 91, 87, 83

She gives a rank of 1 to the lowest score and so on. What will be the difference in the sum of the ranks for the extrovert and introvert groups?

a) 14 b) 0 c) -18 d) -14 e) 18

Questions 23-26 are based on the following study :

23. Suppose a STA 3024 instructor wants to test three kinds of exam formats for her class. These are : (1) only multiple choice (2) only short answer type and (3) a mix of multiple choice and short answer type. Accordingly, she administers type (1) for the first , type (2) for the second and type (3) for the third exam. The scores of 4 randomly chosen students for each of the exams are as follows

Type I/Exam I : 77, 90, 88, 65
Type II/Exam II : 67, 85, 89, 73
Type III/Exam III : 90, 65, 73, 80

Suppose the instructor wants to perform a Kruskal Wallis test to test for any difference between the three testing formats. What should be her alternative hypotheses ?

a) The mean scores corresponding to the three exams are different.
b) The population distribution of the scores for the three testing formats are different.
c) At least one of the mean exam scores is different from the rest.
d) The population distribution of atleast one of the exam scores is different from the rest.
e) The median scores corresponding to the three exams are different.

Extra Credit Question

24. What will be the value of the Kruskal Wallis test statistic?

a) 1.25 b) 12.5 c) -0.125 d) 2.5 e) 0.125

25. Based on the p-value of the above test what can you conclude? *(You can assume that H_0 : Identical population distribution of the scores for the 3 exam types).*

a) Reject H_0 at $\alpha = 0.1$ but not at $\alpha = 0.05$ and $\alpha = 0.01$.
b) Reject H_0 at $\alpha = 0.05$ and $\alpha = 0.1$ but not at $\alpha = 0.01$.
c) Reject H_0 at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.
d) Fail to reject H_0 at $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$.

26. Suppose all the assumptions of ANOVA are satisfied for the above problem. Let p^{kw} and p^a respectively be the p-values corresponding to the Kruskal Wallis and the ANOVA test. Which of the following depicts a plausible relation between the two p-values?

a) $p^{kw} \geq p^a$ b) $p^{kw} \leq p^a$ c) $p^{kw} = p^a$ d) $p^{kw} = 2p^a$ e) $p^{kw} = p^a/2$

Questions 27-30 are based on the following study :

I showed you a recent news where US researchers have claimed that regular meditation practise leads to a significant reduction in death rates, heart attacks and strokes. Suppose an UF mental health physician randomly selects 10 couples. For each pair, she randomly assigns one of them to lifestyle change routines (combination of exercise, biofeedback, stress reduction education etc) and the other one to regular twice a day meditation practice. After 6 months, she asks each of them to rate their overall wellness (a combination of mental and physical well being) on a scale of 1 - 5 (1 denoting excellent and 5 denoting really bad). The data are given below

Couple →	1	2	3	4	5	6	7	8	9	10
Lifestyle	2	2	4	1	3	4	1	2	2	3
Meditation	1	1	3	2	4	1	2	3	1	1

The physician wants to perform a sign test in order to analyze whether regular meditation practice results in a more profound improvement in wellness than lifestyle change routines.

27. Supposing p denotes the population proportion of couples for whom meditation is more beneficial than lifestyle changes, what would be the alternative hypotheses for the sign test?

a) $p \leq 0.5$ b) $p = 0.5$ c) $p \neq 0.5$ d) $p \geq 0.5$

28. What would be the test statistic and the corresponding p-value for the above test?

a) $z = 0.63; p = 0.26$ b) $z = -1.45; p = 0.93$ c) $z = 0.63; p = 0.7357$ d) $z = 1.45; p = 0.07$

29. Suppose the researcher next decides to perform a Wilcoxon signed rank test to be absolutely sure of her conclusions. What will be the test statistic value?

a) 33 b) 37 c) 46 d) 55 e) 55

30. For the above test statistic, the software generates a p-value of 0.032. What should be your conclusion at 5% level of significance?

a) Population median of the difference scores (between lifestyle change and meditation) is negative i.e meditation and lifestyle routines are equally effective in inducing wellness.
b) Population median of the difference scores (between lifestyle change and meditation) is positive i.e meditation is inferior to lifestyle routines in inducing wellness.
c) Population median of the difference scores (between lifestyle change and meditation) is negative i.e meditation is superior to lifestyle routines in inducing wellness.
d) Population median of the difference scores (between lifestyle change and meditation) is positive i.e meditation is superior to lifestyle routines in inducing wellness.