HH/HLST 2300: Statistical Methods in Health Studies
Winter Term Assignment 3
Assigned: Friday February 12, 2021; Due 5PM Friday March 5, 2021
Submit 1 file for Assignment 3: PDF
PDF document name: LASTNAME_FIRSTNAME_WTAssignment3
Submit via eClass

Note1: WT Assignment 3 is worth a total of 43 marks. Therefore, assignments submitted late were deducted 2.15 marks per day (43 * 0.05 = 2.15)
Note2: If you did not use the proper assignment naming convention, 4 marks were deducted.

As noted in my eClass announcement on Nov 5, 2020, you will be deducted marks if you submit a file other than a pdf file and if that file is named incorrectly. **For WT Assignment 3, the deduction is 4 marks for incorrect file name. The submission file type has been set up in eClass such that the only accepted file type is PDF.**

Other reminders that you should take care to ensure while completing your assignment:
- Questions involving a data file must be answered using SPSS
- HLST 2300 rounding rules apply unless otherwise stated
- Screenshots of any hand-written work and SPSS must be of high resolution and be pasted upright (not sideways) so that they can be easily read and graded
- Answers to questions must directly follow the question asked – do not change the order of the questions
- If you fail to include the SPSS output instructed of you, you will receive zero for any subsequent questions that rely on that output

1. The Emergency Department (ED) manager of rural Hospital ABC would like to know if there is an association between triage level and arrival mode. The below contingency table summarizes data recorded in the ED of Hospital ABC last week. Run the appropriate test in SPSS. Copy and paste the relevant output table(s) and report the results. **[7 MARKS]**

	Ambulance	Walk-in
Resuscitation	8	6
Emergency	57	42
Urgent	19	26

triage_level * arrival_mode Crosstabulation

			arrival_mode ambulance	arrival_mode walk-in	Total
triage_level	resuscitation	Count	8	6	14
		Expected Count	7.4	6.6	14.0
		% within triage_level	57.1%	42.9%	100.0%
	emergency	Count	57	42	99
		Expected Count	52.6	46.4	99.0
		% within triage_level	57.6%	42.4%	100.0%
	urgent	Count	19	26	45
		Expected Count	23.9	21.1	45.0
		% within triage_level	42.2%	57.8%	100.0%
Total		Count	84	74	158
		Expected Count	84.0	74.0	158.0
		% within triage_level	53.2%	46.8%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	3.027[a]	2	.220	.237		
Likelihood Ratio	3.028	2	.220	.237		
Fisher-Freeman-Halton Exact Test	3.022			.237		
Linear-by-Linear Association	2.266[b]	1	.132	.169	.085	.036
N of Valid Cases	158					

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.56.

b. The standardized statistic is 1.505.

A χ^2 test of independence was performed to test the null hypothesis that there was no association between triage level and arrival mode. We found no statistically significant association between triage level and arrival mode $\chi^2(2) = 3.03$, p = .237.

Some of the typical errors found in Question 1:
- Did not correctly generate a Crosstab Table of triage level by arrival mode – categories must be stated with labels (resuscitation, emergency, urgent, ambulance, walk-in) and not values. [-1 MARK]
- Observed counts in the crosstab table are different than those shown. [-1 MARK]
- Expected counts in the crosstab table are different than those shown. [-1 MARK]
- Note, if reported results without producing the corresponding tables where these results came from, no marks are earned.

2. National-level hospital data show that all-terrain vehicle, boat, cycling, scooter, ski/snowboard, and snowmobile account for 24.6%, 3.7%, 39.4%, 3.3%, 18.1% and 10.9%, respectively, of sports equipment-related injuries. Data collected from Ontario of 100 sports equipment-related injuries have the following counts:

Sports equipment	All-terrain vehicle	Boat	Cycling	Scooter	Ski/snowboard	Snowmobile
Injuries	22	5	45	4	14	10

Do these data provide sufficient evidence to indicate that the proportions of sports equipment injuries in Ontario differ from the proportions for the national-level population? Run the appropriate test in SPSS. Copy and paste the relevant output table(s) and report the results. **[7 MARKS]**

sports_equip_inj

	Observed N	Expected N	Residual
All-terrain vehicle	22	24.6	-2.6
Boat	5	3.7	1.3
Cycling	45	39.4	5.6
Scooter	4	3.3	.7
Ski/snowboard	14	18.1	-4.1
Snowmobile	10	10.9	-.9
Total	100		

Test Statistics

	sports_equip_inj
Chi-Square	2.679[a]
df	5
Asymp. Sig.	.749
Exact Sig.	.752
Point Probability	.000

a. 2 cells (33.3%) have expected frequencies less than 5. The minimum expected cell frequency is 3.3.

A χ^2 goodness-of-fit test was performed to test the null hypothesis that all-terrain vehicle, boat, cycling, scooter, ski/snowboard, and snowmobile accounted for 24.6%, 3.7%, 39.4%, 3.3%, 18.1% and 10.9%, respectively, of sports equipment injuries in Ontario as they do in national-level data. The results were not statistically significant and we fail reject the null hypothesis, $\chi^2(5) = 2.68$, p = .752.

Some of the typical errors found in Question 2:
- Did not generate an Observed and Expected Table of sports equipment injury categories – categories must be stated with labels and not values. [-1 MARK]
- Observed counts are different than those shown. [-1 MARK]
- Expected counts are different than those shown. [-1 MARK]
- Note, if reported results without producing the corresponding tables where these results came from, no marks are earned.
- Did not generate Chi Square Test Table or values in the table are different than those shown. [-1 MARK]
- Did not correctly report chi-square value, including degrees of freedom. [-1 MARK]
- Did not correctly report p-value. [-1 MARK]
- Did not report that result was not statistically significant. [-1 MARK]

3. We have implemented a pain intervention and collected pain data (0 = no pain; 1 = pain) on 50 adults measured at 4 intervals (pre-intervention, post-7days, post-30days and post-90 days), Excel file: 2300WTassignment3.xls; worksheet: pain. Are there significant differences in the proportion of pain-experiencing adults across the four time periods? Run the appropriate test in SPSS. Copy and paste the relevant output table(s) and report the results. **[15 MARKS]**

Cochran Test

Frequencies

	Value	
	0	1
pain_pre	20	30
pain_post7days	27	23
pain_post30days	27	23
pain_post90days	34	16

Test Statistics

N	50
Cochran's Q	12.511[a]
df	3
Asymp. Sig.	.006
Exact Sig.	.005
Point Probability	.001

a. 1 is treated as a success.

pain_pre & pain_post7days

pain_pre	pain_post7days	
	no pain	pain
no pain	15	5
pain	12	18

pain_pre & pain_post30days

pain_pre	pain_post30days	
	no pain	pain
no pain	15	5
pain	12	18

pain_pre & pain_post90days

pain_pre	pain_post90days	
	no pain	pain
no pain	18	2
pain	16	14

pain_post7days & pain_post30days

pain_post7days	pain_post30days	
	no pain	pain
no pain	19	8
pain	8	15

pain_post7days & pain_post90days

pain_post7days	pain_post90days	
	no pain	pain
no pain	22	5
pain	12	11

pain_post30days & pain_post90days

pain_post30days	pain_post90days	
	no pain	pain
no pain	26	1
pain	8	15

Test Statistics[a]

	N	Exact Sig. (2-tailed)	Exact Sig. (1-tailed)	Point Probability	
pain_pre & pain_post7days	50	.143[b]	.072	.047	px = k*ps = 6(.143) = .858
pain_pre & pain_post30days	50	.143[b]	.072	.047	px = k*ps = 6(.143) = .858
pain_pre & pain_post90days	50	.001[b]	.001	.001	px = k*ps = 6(.001) = .006
pain_post7days & pain_post30days	50	1.000[b]	.598	.196	px = k*ps = 6(1.000) = 6.000
pain_post7days & pain_post90days	50	.143[b]	.072	.047	px = k*ps = 6(.143) = .858
pain_post30days & pain_post90days	50	.039[b]	.020	.018	px = k*ps = 6(.039) = .234

a. McNemar Test

b. Binomial distribution used.

Using Cochran's Q test, we found that the proportion of pain-experiencing adults was statistically significantly different across the four intervals, Q(3) = 12.51, p = .005. The McNemar test, with significant levels corrected for multiple testing (Bonferroni correction) showed that significant decreases in pain occurred from pre-intervention to post-90 days (p = .006). There were no significant differences between pre-intervention and post-7 days (p = .858), pre-intervention and post-30 days (p = .858), post-7 days and post-30 days (p = 1.00), post-7 days and post-90 days (p = .858), or between post-30 days and post-90 days (p = .234)

Some of the typical errors found in Question 3:
- Note, if reported results without producing the corresponding tables where these results came from, no marks are earned.
- Did not generate Cochrane Frequencies table with the four timepoints of pain measurement [-1 MARK]
- Did not generate Cochrane Test Statistics table or values in the table are different than those shown. [-1 MARK]
- Did not generate all six McNemar contingency tables. [-1 MARK]
- Did not generate McNemar Test Statistics table or values in the table are different than those shown. [-1 MARK]
- Did not correctly report Cochrane Q value, including degrees of freedom. [-1 MARK]
- Did not correctly report p-value. [-1 MARK]
- Did not report that result was statistically significant. [-1 MARK]
- Did not correctly report Bonferroni-corrected p-values for the six pairwise comparisons. [-1 MARK per incorrect p-value]. Note 1: if did not show calculations for Bonferroni-corrected p-values, no marks were earned. Note 2: for Bonferroni-corrected p-value calculations where p > 1, p-value is stated as p = 1.00 since by definition the maximum value of a probability is 1.
- Did not report that pre-intervention to post-90 days was statistically significant. [-1 MARK]
- Did not report that the five remaining pairwise combinations were not statistically significant. [-1 MARK]

4. Ontario has collected data for COVID-19 cases in long-term care (LTC) homes (Excel file: 2300WTassignment3.xls; worksheet: covid). The worksheet includes the daily count (for a full month) of the number of confirmed active resident cases of COVID-19 (Confirmed_Active_LTC_Resident_Cases) and the number of confirmed active home care worker cases (Confirmed_Active_LTC_HCW_Cases). Is there a positive correlation between confirmed active resident cases of COVID-19 and confirmed active home care worker cases? Test this hypothesis by running the appropriate test in SPSS. Copy and paste the relevant SPSS output table(s) and report the results. **[8 MARKS]**

Statistics

		Confirmed_Active_LTC_Resident_Cases	Confirmed_Active_LTC_HCW_Cases
N	Valid	31	31
	Missing	0	0

Both confirmed active resident cases of COVID-19 and confirmed active home care worker cases of COVID-19 are scalar variables and we have a large sample size (n≥ 30), so Pearson's r may be appropriate so long as a linear trend exists between the two variables.
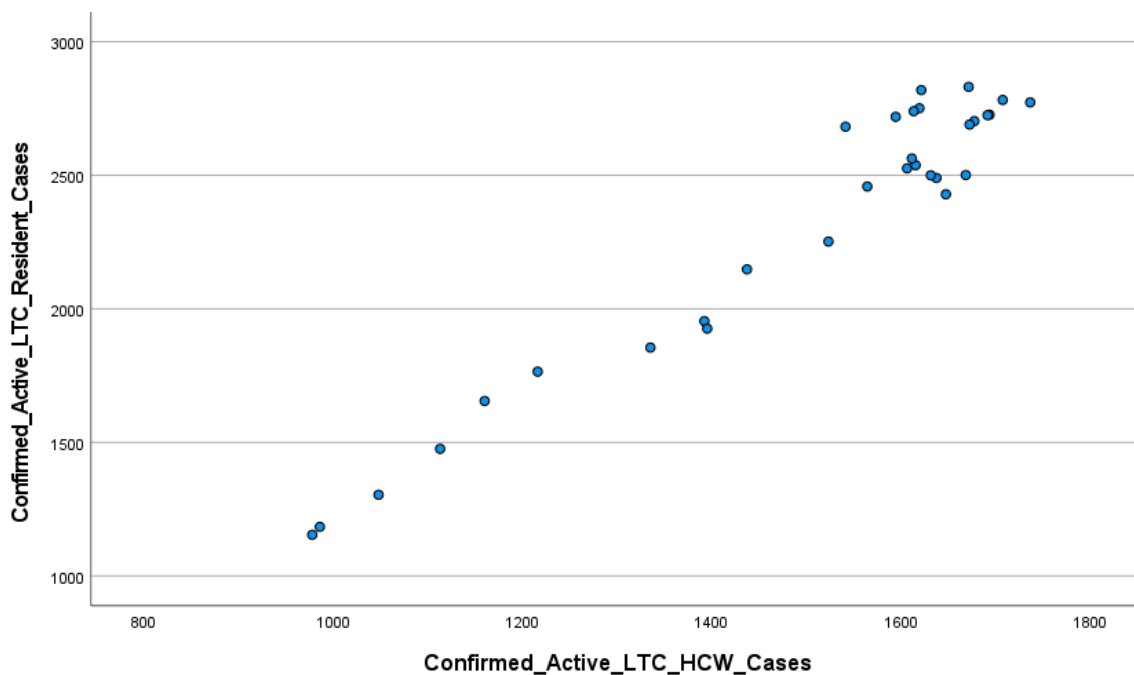
The scatterplot shows a positive linear trend.

Correlations

		Confirmed_Active_LTC_Resident_Cases	Confirmed_Active_LTC_HCW_Cases
Confirmed_Active_LTC_Resident_Cases	Pearson Correlation	1	.973**
	Sig. (1-tailed)		.000
	N	31	31
Confirmed_Active_LTC_HCW_Cases	Pearson Correlation	.973**	1
	Sig. (1-tailed)	.000	
	N	31	31

**. Correlation is significant at the 0.01 level (1-tailed).

The linear relationship between confirmed active resident cases of COVID-19 and confirmed active home care worker cases of COVID-19 was measured using Pearson's correlation coefficient. There was a strong, positive relationship which was statistically significant, $r = 0.97$, $p < .001$.

Some of the typical errors found in Question 4:
- Did not generate a frequency table of confirmed active resident cases of COVID-19 and confirmed active home care worker cases (both scalar variables) confirming large samples size to justify use of Pearson r. Note confirming large sample size must happen *before* conducting Pearson's r. [-1 MARK]
- Did not produce a scatterplot of confirmed active resident cases of COVID-19 and confirmed active home care worker cases confirming linear trend. [-1 MARK]
- Did not copy and paste the Pearson r Correlations Table or values in the table are different than those shown (eg incorrect to generate 2-tailed p-value since our hypothesis had a direction - positive correlation). [-1 MARK]
- Note, if reported results without producing the corresponding table where these results came from, no marks are earned.
- Did not report that relationship was strong in strength. [-1 MARK]
- Did not report that relationship was positive in direction. [-1 MARK]
- Did not report r value or value reported is incorrect. [-1 MARK]
- Did not report p-value or value reported is incorrect. [-1 MARK]
- Did not report that result was statistically significant. [-1 MARK]

5. A researcher has collected data for 158 adult (age ≥ 18 yrs) patients arriving via the Emergency Department (ED) and admitted as an inpatient to Hospital ABC (Excel file: 2300WTassignment3.xls; worksheet: riw). The data includes the unique patient identifier, comorbidity level (levels range from 0 – 4 where level 0 represents no significant comorbidity and level 4 represents the group with the largest number of comorbidities) and resource intensity weight (RIW) which is a proxy for hospital resource use. Is there a positive correlation between comorbidity level and RIW? Test this hypothesis by running the appropriate test in SPSS. Copy and paste the relevant SPSS output table(s) and report the results. **[6 MARKS]**

Correlations

			Comorbidity level	RIW
Spearman's rho	Comorbidity level	Correlation Coefficient	1.000	.665**
		Sig. (1-tailed)	.	.000
		N	158	158
	RIW	Correlation Coefficient	.665**	1.000
		Sig. (1-tailed)	.000	.
		N	158	158

**. Correlation is significant at the 0.01 level (1-tailed).

Spearman's correlation was used to examine the association between comorbidity level and RIW since comorbidity level is an ordinal variable. There was a moderate, positive relationship that was statistically significant, $r_s = .67$, $p < .001$.

Some of the typical errors found in Question 5:
- Did not copy and paste the Spearman rho Correlations Table or values in the table are different than those shown (eg incorrect to generate 2-tailed p-value since our hypothesis had a direction - positive correlation). [-1 MARK]
- Note, if reported results without producing the corresponding table where these results came from, no marks are earned.
- Did not report that relationship was moderate in strength. [-1 MARK]
- Did not report that relationship was positive in direction. [-1 MARK]
- Did not report rho value or value reported is incorrect. [-1 MARK]
- Did not report p-value or value reported is incorrect. [-1 MARK]
- Did not report that result was statistically significant. [-1 MARK]

School of Health Policy and Management

Assignment Attachment Form

Student Name:

Student Number:

Course Code:

Assignment Title:

Due Date:

Tutorial Leader (if applicable):
Please check each box after reading, to acknowledge agreement with each statement.

☐ I have read and understand the Senate Policy on Academic Honesty found on website at the following York Secretariat website on Academic Honesty.

☐ I have read and understood the assignment submission described in the course outline (syllabus)

☐ I have read and understood the criteria used for assessment in this assignment

☐ I have read and understood and followed the referencing guidelines required for assignments submitted at York University

☐ This assignment is entirely my own work, except where I have given documented references to work of others

☐ This assignment or substantial parts of it has not previously been submitted for assessment in any formal course of study, unless acknowledged in the assignment and previously agreed to by my Tutorial Leader and Course Director

☐ I understand that this assignment may undergo electronic detection for plagiarism and a copy of the assignment may be retained on the database and used to make comparisons with other assignments in the future

Signature: _____ Date: _____