Université Libanaise
Faculté d'Information

الجامعة اللبنانية
كلية الاعلام

العام الجامعي 2022 (دورة ثانية)

المرحلة: إجازة	**المادة: statistics 2**
السنة: 2022	**المدة: ساعة ونصف**
الاختصاص: علوم البيان	**اسم الأستاذ: د. حسين نصار**

Section 1: answer the following multiple choice questions:

1- The least squares regression line y = 8 − 9x predicts that, when x = 5, the value of y is:

A) −45
B) −37
C) 37
D) 45
E) 53

2- A least squares regression line of the form y = a + b x is fitted to the data set below.

x	25	15	10	5
y	10	10	15	25

The equation of the line is:
A) y = −0.69 + 24.4x
B) y = 24.4 − 0.69x
C) y = 24.4 + 0.69x
D) y = 28.7 − x
E) y = 28.7 + x

3- Weight (in kg) can be predicted from height (in cm) from the least squares regression line:

Weight = −96 + 0.95 × Height, with r = 0.79

A) decreases by 96 kg for each one centimetre increase in height
B) increases by 96 kg for each one centimetre increase in height
C) decreases by 0.79 kg for each one centimetre increase in height
D) decreases by 0.95 kg for each one centimetre increase in height
E) increases by 0.95 kg for each one centimetre increase in height

4- A person of height 179 cm weighs 82 kg. If the regression equation in previous part is used to predict their weight, then the residual will be closest to:

A) −8 kg
B) 3 kg
C) 8 kg
D) 9 kg
E) 74 kg

5- In the least squares regression line y = −1.2 + 0.52x :

A the y-intercept = −0.52 and slope = −1.2
B the y-intercept = 0 and slope = −1.2
C the y-intercept = 0.52 and slope = −1.2
D the y-intercept = −1.2 and slope = 0.52
E the y-intercept 12 and slope = -0.52

Section 2 : fill in the blanks with the right word(s)

In an investigation of the relationship between the hours of sunshine (per year) and days of rain (per year) for 25 cities, the least squares regression line was found to be:
Hours of sunshine = 2847 − 6.88 × Days of rain, with $r^2 = 0.4838$
Use this information to complete the following sentences.
a. In this regression equation, the independent variable is _____ .
b. The slope is _____ and the intercept is _____ .
c. The regression equation predicts that a city that has 120 days of rain per year will have _____ hours of sunshine per year.
d. The slope of the regression line predicts that the hours of sunshine per year will _____by _____ hours for each additional day of rain.
e. r = _____ , correct to three decimal places.
f. _____% of the variation in sunshine hours can be explained by the variation in_____ .

Section 3

Four brands of flashlight batteries are to be compared by testing each brand in five flashlights. Twenty flashlights are randomly selected and divided randomly into four groups of five flashlights each. Then each group of flashlights uses a different brand of battery. The lifetimes of the batteries, to the nearest hour, are as follows.

Brand *A*	Brand *B*	Brand *C*	Brand *D*
42	28	24	20
30	36	36	32
39	31	28	38
28	32	28	28
29	27	33	25

Source	df	SS	MS = SS/df	F-statistic	*p*-value
Treatments	3	68.2	22.7333	0.7393	*p*-value > 0.10
Error	16	492.0	30.75		
Total	19	560.2			

Preliminary data analyses indicate that the independent samples come from normal populations with equal standard deviations. At the 5% significance level, does there appear to be a difference in mean lifetime among the four brands of batteries? (use the 6 steps approach)

Section 4 (40 points)

Part I

A researcher in criminology is curious about the relation between police patrols in his city and the crime rate. Over a period of one year, he collected data from 34 housing complexes in the city. For each complex he recoded whether there was a crime committed in the vicinity and whether police patrol is high or low in that area. The following tables are the results from SPSS output file:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PolicePatrol * Crime	34	100.0%	0	0.0%	34	100.0%

PolicePatrol * Crime Crosstabulation

			Crime		
			Yes	No	Total
PolicePatrol	High	Count	5	14	19
		Expected Count	8.4	10.6	19.0
		% within PolicePatrol	26.3%	73.7%	100.0%
	Low	Count	10	5	15
		Expected Count	6.6	8.4	15.0
		% within PolicePatrol	66.7%	33.3%	100.0%
Total		Count	15	19	34
		Expected Count	15.0	19.0	34.0
		% within PolicePatrol	44.1%	55.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.536[a]	1	.019		
Continuity Correction[b]	4.020	1	.045		
Likelihood Ratio	5.666	1	.017		
Fisher's Exact Test				.036	.022
Linear-by-Linear Association	5.373	1	.020		
N of Valid Cases	34				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.62.

b. Computed only for a 2x2 table

Answer the following questions:

1) Why is chi-sqaure test considered appropriate in this case?
2) State the null and alternative hypothesis
3) What can you conclude from the tables regarding researcher's inquiry?

Part II

The researcher has doubts that the results that he arrived might not be totally valid, this is because he suspects that another variable namely the "location" of the housing complex (measured as a either in downtown or suburbs) may have an impact on the analysis. He conducted another test taking into account this idea and the following were the results:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PolicePatrol * Crime * Location	34	100.0%	0	0.0%	34	100.0%

PolicePatrol * Crime * Location Crosstabulation

Location			Crime		Total
			Yes	No	
Downtown	PolicePatrol	High	Count	3	6	9
			Expected Count	5.1	3.9	9.0
			% within PolicePatrol	33.3%	66.7%	100.0%

		Low	Count	5	0	5
			Expected Count	2.9	2.1	5.0
			% within PolicePatrol	100.0%	0.0%	100.0%
	Total		Count	8	6	14
			Expected Count	8.0	6.0	14.0
			% within PolicePatrol	57.1%	42.9%	100.0%
Suburbs	PolicePatrol	High	Count	2	8	10
			Expected Count	3.5	6.5	10.0
			% within PolicePatrol	20.0%	80.0%	100.0%
		Low	Count	5	5	10
			Expected Count	3.5	6.5	10.0
			% within PolicePatrol	50.0%	50.0%	100.0%
	Total		Count	7	13	20
			Expected Count	7.0	13.0	20.0
			% within PolicePatrol	35.0%	65.0%	100.0%
Total	PolicePatrol	High	Count	5	14	19
			Expected Count	8.4	10.6	19.0
			% within PolicePatrol	26.3%	73.7%	100.0%
		Low	Count	10	5	15
			Expected Count	6.6	8.4	15.0
			% within PolicePatrol	66.7%	33.3%	100.0%
	Total		Count	15	19	34
			Expected Count	15.0	19.0	34.0
			% within PolicePatrol	44.1%	55.9%	100.0%

Chi-Square Tests

Location		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Downtown	Pearson Chi-Square	5.833[c]	1	.016		
	Continuity Correction[b]	3.429	1	.064		
	Likelihood Ratio	7.664	1	.006		
	Fisher's Exact Test				.031	.028

	Linear-by-Linear Association	5.417	1	.020		
	N of Valid Cases	14				
Suburbs	Pearson Chi-Square	1.978[d]	1	.160		
	Continuity Correction[b]	.879	1	.348		
	Likelihood Ratio	2.027	1	.155		
	Fisher's Exact Test				.350	.175
	Linear-by-Linear Association	1.879	1	.170		
	N of Valid Cases	20				
Total	Pearson Chi-Square	5.536[a]	1	.019		
	Continuity Correction[b]	4.020	1	.045		
	Likelihood Ratio	5.666	1	.017		
	Fisher's Exact Test				.036	.022
	Linear-by-Linear Association	5.373	1	.020		
	N of Valid Cases	34				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.62.

b. Computed only for a 2x2 table

c. 3 cells (75.0%) have expected count less than 5. The minimum expected count is 2.14.

d. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 3.50.

1) What do we call this new test and why is it helpful ?
2) What conclusions can you make out of the tables above regarding the association between police patrol and crime ?