



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

Answer ALL the following questions:

**Question 1 (30 points):**

A researcher would like to investigate the impact of studying sound environment on the performance of students. He considered three types of sounds: classical music, white noise, and no sound at all.

The following is the data collected from three samples from the three sound environments:

	Scores of the three different types of class environments									
Classical Music	7	9	5	8	6	8	6	10	7	4
White Noise	4	3	6	2	7	5	5	4	1	3
No sound	5	1	3	5	3	4	6	5	7	3

**Summary**

	Summary of Data			
	Count	Sum	Average	Variance
Classical Music	10	70	7	3.3333
White Noise	10	40	4	3.3333
No sound	10	43	4.3	3.3444

- The researcher has decided to conduct an ANOVA testing to figure out if there is differences between the different class environments. Why do you think ANOVA is a good test for this problem?
- Use the 6-steps hypothesis testing approach and the table below to make a conclusion regarding the problem the researcher is trying to solve.

0 Modified

ANOVA							
Source of variation	SS	DF	MS	F	p-value	F crit	
Between Groups	54.6	2	27.3	8.18	0.001669	3.3541308	
Within Groups	90.1	27	3.337038				
Total	144.7	29					

c) Using the 95% confidence interval and the following formula find out which class environment might be different from the other? What is the overall conclusion that you can make?

$$(\bar{x}_N - \bar{x}_B) \pm t \sqrt{\text{MSE} \left( \frac{1}{n_N} + \frac{1}{n_B} \right)}$$

**Question 2 (40 points):**

A farmer who owns a substantial land is interested in improving the crops. He is puzzled by the fact that the yields are varying from one area to another within his vast land, which is scattered within the whole country. He consulted a researcher to know the factors that might potentially have an impact on the crops. The researcher took a sample from 35 different areas and installed sensors to monitor the amount of sun light, amount of water and amount of nutrients that the crops are receiving.

The following is a sample of the data:

Light (hours daily)	Water in inches	Nutrients (milligrams)	Crops (grams)
5.3	6	300	600
6.5	7.5	700	850
7.1	7.4	800	855
7.1	8	850	800
8.7	5.7	450	750
9.3	8.4	566	790
7.3	7	450	600
7.6	9	980	900
8.1	9	489	700

The following were the results:

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0.832589								
R Square	0.693204								
Adjusted R Square	0.663515								
Standard Error	77.6662								
Observations	35								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	3	422511.1	140837	23.34816121	4.27E-08				
Residual	31	186993.2	6032.039						
Total	34	609504.3							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat.</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
Intercept	66.6082	109.5902	0.607793	0.547751771	-156.9025458	290.1189	-156.903	290.11895	
Light	23.86261	13.82014	1.726655	0.094183892	-4.32374218	52.04897	-4.32374	52.04896593	
Water in inch	33.94489	13.02871	2.605391	0.013972412	7.372653882	60.51713	7.372654	60.51713215	
Nutrients	0.387196	0.067226	5.759645	2.44E-06	0.250088447	0.524304	0.250088	0.524303909	

- What statistical method did the researcher use? Explain what the method does and why it might be a good or a bad candidate to solve the problem that the land owner faces
- Explain what is meant by Multiple R, R Square, Adjusted R Square and Standard Error.
- Explain why the ANOVA table was used.
- What conclusions can you derive from the ANOVA table (use the 6 steps process).
- Explain why the last table was used.
- What conclusions can you draw from the last (third) table regarding each of the independent variables? If the experiment were to be repeated which independent variable would you suggest to remove and why?
- Assuming the model in the tables above is true. How much crops you predict to obtain in a land with 7 hours of daily sunlight, 8 inches of water and 600 milligrams of nutrients?

**Question 3 (30 points):**

A researcher is interested in knowing whether the neighborhoods within a city has an impact on the resident's pursuit of university education. To conduct this study he collected information from two neighborhoods (A and B) and asked each respondents about their education attained. After that he performed a chi-square test and this was the result:

Case Processing Summary					
Valid		Cases			
		Missing		Total	
N	Percent	N	Percent	N	Percent

Neighborhood * Education	165	100.0%	0	0.0%	165	100.0%
--------------------------	-----	--------	---	------	-----	--------

Neighborhood \* Education Crosstabulation

Count

		Education		Total
		High school	University	
Neighborhood	A	42	28	70
	B	31	64	95
Total		73	92	165

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	12.238 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	11.153	1	.001		
Likelihood Ratio	12.332	1	.000		
Fisher's Exact Test				.001	.000
Linear-by-Linear Association	12.163	1	.000		
N of Valid Cases	165				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.97.

b. Computed only for a 2x2 table

- Why did the researcher use chi-square test? Give all the reasons.
- Interpret the results in the table above indicating the six steps process of hypothesis testing.

The researcher thinks that the economic status of the individual is causing the discrepancy in education. To check if these thought were valid, he conducted another test. The tables below display the results of this test:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Neighborhood * Education * Economic	165	100.0%	0	0.0%	165	100.0%

Neighborhood \* Education \* Economic Crosstabulation

Count

Economic			Education		Total
			Highschool	Unvrsly	
Low	Neighborhood	A	26	12	38
		B	12	5	17
	Total			38	17
Medium	Neighborhood	A	10	6	16
		B	7	19	26
	Total			17	25
High	Neighborhood	A	6	10	16
		B	12	40	52
	Total			18	50
Total	Neighborhood	A	42	28	70
		B	31	64	95
	Total			73	92

#### Chi-Square Tests

Economic		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Low	Pearson Chi-Square	.026*	1	.872	1.000	.567
	Continuity Correction <sup>a</sup>	.000	1	1.000		
	Likelihood Ratio	.026	1	.872		
	Fisher's Exact Test					
	Linear-by-Linear Association	.025	1	.873		
	N of Valid Cases	55				
Medium	Pearson Chi-Square	5.203 <sup>a</sup>	1	.023	.029	.025
	Continuity Correction <sup>a</sup>	3.832	1	.050		
	Likelihood Ratio	5.232	1	.022		
	Fisher's Exact Test					
	Linear-by-Linear Association	5.080	1	.024		
	N of Valid Cases	42				
High	Pearson Chi-Square	1.308 <sup>a</sup>	1	.253	.332	.204
	Continuity Correction <sup>a</sup>	.672	1	.412		
	Likelihood Ratio	1.246	1	.264		
	Fisher's Exact Test					
	Linear-by-Linear Association	1.288	1	.256		
	N of Valid Cases	68				
Total	Pearson Chi-Square	12.238 <sup>a</sup>	1	.000	.001	.000
	Continuity Correction <sup>a</sup>	11.153	1	.001		
	Likelihood Ratio	12.332	1	.000		
	Fisher's Exact Test					

Linear-by-Linear Association	12.163	1	.000		
N of Valid Cases	165				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.97.

b. Computed only for a 2x2 table

c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.25.

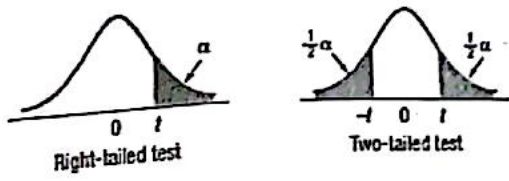
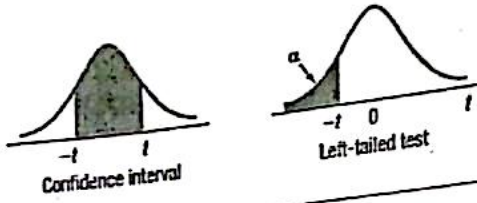
d. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.48.

e. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.24.

c) What type of test did the researcher make and why is this test important and what do we call the variable "economic status"?

d) Discuss and interpret the last table above highlighting what it implies for the overall conclusions of the conducted research.

# B.2 Student's *t* Distribution



df	Confidence Intervals, <i>c</i>					
	Level of Significance for One-Tailed Test, $\alpha$					
	Level of Significance for Two-Tailed Test, $\alpha$					
	0.10	0.05	0.025	0.01	0.005	0.0005
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.898	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.728	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.509	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
31	1.309	1.696	2.040	2.453	2.744	3.633
32	1.309	1.694	2.037	2.449	2.738	3.622
33	1.308	1.692	2.035	2.445	2.733	3.611
34	1.307	1.691	2.032	2.441	2.728	3.601
35	1.306	1.690	2.030	2.438	2.724	3.591

df	Confidence Intervals, <i>c</i>					
	Level of Significance for One-Tailed Test, $\alpha$					
	Level of Significance for Two-Tailed Test, $\alpha$					
	0.10	0.05	0.025	0.01	0.005	0.0005
36	1.305	1.688	2.028	2.434	2.719	3.582
37	1.305	1.687	2.026	2.431	2.715	3.574
38	1.304	1.686	2.024	2.429	2.712	3.566
39	1.304	1.685	2.023	2.426	2.708	3.558
40	1.303	1.684	2.021	2.423	2.704	3.551
41	1.303	1.683	2.020	2.421	2.701	3.544
42	1.302	1.682	2.018	2.418	2.699	3.538
43	1.302	1.681	2.017	2.416	2.695	3.532
44	1.301	1.680	2.015	2.414	2.692	3.526
45	1.301	1.679	2.014	2.412	2.690	3.520
46	1.300	1.679	2.013	2.410	2.687	3.515
47	1.300	1.678	2.012	2.408	2.685	3.510
48	1.299	1.677	2.011	2.407	2.682	3.505
49	1.299	1.677	2.010	2.405	2.680	3.500
50	1.299	1.676	2.009	2.403	2.678	3.496
51	1.298	1.675	2.008	2.402	2.676	3.492
52	1.298	1.675	2.007	2.400	2.674	3.488
53	1.298	1.674	2.006	2.399	2.672	3.484
54	1.297	1.674	2.005	2.397	2.670	3.480
55	1.297	1.673	2.004	2.396	2.668	3.476
56	1.297	1.673	2.003	2.395	2.667	3.473
57	1.297	1.672	2.002	2.394	2.665	3.470
58	1.296	1.672	2.002	2.392	2.663	3.466
59	1.296	1.671	2.001	2.391	2.662	3.463
60	1.296	1.671	2.000	2.390	2.660	3.460
61	1.296	1.670	2.000	2.389	2.659	3.457
62	1.295	1.670	1.999	2.388	2.657	3.454
63	1.295	1.669	1.998	2.387	2.656	3.452
64	1.295	1.669	1.998	2.386	2.655	3.449
65	1.295	1.669	1.997	2.385	2.654	3.447
66	1.295	1.668	1.997	2.384	2.652	3.444
67	1.294	1.668	1.996	2.383	2.651	3.442
68	1.294	1.668	1.995	2.382	2.650	3.439
69	1.294	1.667	1.995	2.382	2.649	3.437
70	1.294	1.667	1.994	2.381	2.648	3.435

(continued)