

Assignment (in lieu of Minor II): SBL701 Biometry**Max Marks: 30****Submission deadline:** 1700 hrs on 14th November, 2022.**Submission details (deviations are not acceptable):**

- I. Only soft copy submissions on the email id amittal@bioschool.iitd.ac.in will be accepted (i.e. no hard copy). No late submission will be accepted (regardless of server problems etc.). To be safe, submit well in advance.
- II. Submissions should be compiled in form of one Adobe pdf file. Multiple files or files in any other software are not acceptable. One way to do this is "print" all your files as "adobe pdf" rather than sending them to a printer.
- III. Use Microsoft Equation 3.0 or Equation Editor or Math Type in MS word if you need to type any equations.
- IV. Solutions should contain each and every step followed for calculations, all raw data (properly tabulated), tables, graphs, programming codes (MATLAB), code output (in form of numbers in command window or figures).
- V. Copying is not acceptable. While you are free to consult anyone, the solutions, interpretations, answers should all be in your own language. Any material that is found to be copied from another source will automatically lead to a "zero" out of 30. Any material found to be identical in different submissions will also lead to "zero" out of 30 for all the submissions appearing identical.

Questions

1. Some mammalian cells were grown in culture medium under five (05) different conditions, in ten (10) 35 mm diameter dishes for each condition (lets call each group of experiments with 10 dishes each as a "Set"). After a fixed time, cells were counted in each of the 10 dishes, for each of the conditions, and the following data was obtained:

Cell Count per ml in Condition (x 10 ⁻⁶)	1	2	3	4	5
Set 1	6.2	5.8	5.7	7.5	5.8
Set 2	6.6	5.9	5.8	6.7	6.1
Set 3	6.5	5.8	6	7	5.6
Set 4	6.3	6.1	5.9	7.5	5.8
Set 5	6.4	5.7	6.2	6.5	5.7
Set 6	6.2	5.6	6	7.1	5.6
Set 7	6.5	5.8	6	6.7	6.1
Set 8	6.5	5.7	5.7	6.7	6
Set 9	6.2	5.7	5.9	7.6	5.7
Set 10	6.7	5.9	6.1	6.8	5.8

- (a) How many t-tests will you need to do find the most suitable condition(s) for growth?
- (b) First perform the t-tests by doing calculations yourself (i.e. calculate the t-statistic manually and find the applicable t-crit values from any source). Next, perform the same tests using statistical toolbox of your choice (e.g. MS excel, MATLAB, SAS). Compare the results with your calculations (i.e. compare results seen from first and second approaches) and comment on the findings.
- (c) If the probability of a type 1 error is 5% for each of the t-tests, what is the probability of no type 1 errors among your total t-tests, assuming all the t-tests are independent (which is not true)?
- (d) Do the five different conditions have similar or different effect on cell growth? The critical value for the test statistic = 6.
- (e) Which condition(s) is/are the most suitable for obtaining the highest number of cells?
- (f) What is the probability of all your sets being equally representative of the conditions?
- (g) What assumption(s) did you use for answering (e)?

$$(0.5+5+0.5+2.5+1+1+0.5 = \underline{11})$$

2. Using simulated data show how the p-values in t-tests vary as a function of differences in two sample data sets. How are these results dependent on sample sizes of the data sets?

(03)

SBL701 Biometry: Testing for interactions using ANOVA

3. We wish to check whether certain combinations of temperature and pH could induce synthesis of heat shock proteins. While there is substantial literature showing molecular weight of these proteins using SDS-PAGE as a diagnostic test for their expression, we believe that (a) the data from these gels do not provide proper quantitative information on varying amounts of proteins (especially due to parameters like sample loading in gels and variations in total protein content per cell), and (b) simply looking at bands corresponding to known molecular weights does not confirm varying expression of these proteins in different samples due to the presence of several similar sized proteins in cell culture extracts. Thus, either we can perform more rigorous quantitative western blots (extremely tedious and still suffering from the (a) limitation due to the difficulty in quantifying readouts resulting from use of antibodies), or, we can look at transcription of genes of heat shock proteins. We decide to do the latter, and carry out densitometric measurements of the change in HSP70 mRNA from a physiological baseline state (standard denoted by std) after cells are held at a given temperature and pH for 2 hrs. Because the transcription data is known to be fairly noisy, we do 10 incubations for each combination of pH and temperature. The data obtained for different types of cells is given below:

Values = $[\text{HSP70 mRNA}]_{\text{cond}} - [\text{HSP70 mRNA}]_{\text{std}}$

Type A

	pH		
Temperature	7.2	7.4	7.6
20	8.6164	4.9751	11.0385
	13.0908	2.2572	14.4758
	3.9662	9.4783	10.0444
	6.9221	3.8621	12.3725
	3.2120	6.9688	9.0507
	0.0242	6.3843	10.1231
	10.1981	5.3911	8.2530
	12.2379	2.0468	9.0501
	5.8252	-1.0019	11.4290
	0.4118	-3.8979	8.6281
30	3.6028	16.0523	14.4630
	12.6908	4.2101	13.9744
	13.1973	-2.8490	7.5078
	11.8734	4.2141	9.3574
	2.9464	2.5869	12.7612
	5.4810	4.4355	11.7407
	9.2247	0.1556	16.7839
	7.2817	-2.7806	7.7679
	5.1267	-0.5223	11.2920
	14.6486	10.9820	15.5478
38	-2.8466	6.2272	20.6849
	7.5565	0.6653	23.0745
	0.7424	-0.1211	18.7124
	6.7195	6.8996	23.9130
	4.3531	14.8993	18.4913
	2.0133	-0.8228	22.4014
	10.3028	5.5409	16.1357
	13.4245	5.5447	17.7635
	13.8453	4.0735	15.7731
16.1136	0.4998	15.5492	

Type B

	pH		
Temperature	7.2	7.4	7.6
20	7.8827	4.8936	9.005
	12.3571	2.1757	12.4378
	3.2325	9.3968	8.0064
	6.1884	3.7806	10.3345
	2.4783	6.8873	7.0127
	-0.7095	6.3028	8.0851
	9.4644	5.3096	6.2150
	11.5042	1.9653	7.0121
	5.0915	-1.0834	9.3910
30	-0.3219	-3.9794	6.5901
	1.9519	15.8688	9.8774
	11.0399	4.0266	9.3888
	11.5464	-3.0325	2.9222
	10.2225	4.0306	4.7718
	1.2955	2.4034	8.1756
	3.8301	4.2520	7.1551
	7.5738	-0.0279	12.1983
	5.6308	-2.9641	3.1823
38	3.4758	-0.7058	6.7064
	12.9977	10.7985	10.9622
	-5.4952	5.9329	13.3276
	4.9079	0.3710	15.7172
	-1.9062	-0.4154	11.3551
	4.0709	6.6053	16.5557
	1.7045	14.6050	11.1340
	-0.6353	-1.1171	15.0441
	7.6542	5.2466	8.7784
10.7759	5.2504	10.4062	
11.1967	3.7792	8.4158	
13.4650	0.2055	8.1919	

- (a) To analyze the data, first we just treat the 9 combinations of temperature and pH as if they were just 9 different treatments for a one way ANOVA. What can be inferred about the synthesis of heat shock proteins in the two cell types? Note: Generate the ANOVA tables for the two cell types in the format given below:

Source	df	Sum of Squares	Mean Square	F-value
Treatment				
Error (= w/in)				
Total				

- (b) Now assume there are “nt” different temperatures and “np” different pH values. Let the k^{th} measurement of mRNA at the t^{th} temperature and p^{th} pH be X_{tpk} . Note that each treatment has the same number, “n”, of measurements (i.e. replicates) in our example. Now to partition the treatment sum of squares into components due to temperatures, pH values, and interactions, first let us develop the formalism for $SS_{\text{treat}} (= SS_{b/w})$ for each treatment: (4)

Treating temperature as a single treatment, regardless of pH:

$$SS_{temp} = \sum_{t=1}^{nt} \sum_{p=1}^{np} \sum_{k=1}^n (\bar{x}_t - \bar{\bar{x}})^2$$

where
$$\bar{x}_t = \frac{\sum_{p=1}^{np} \sum_{k=1}^n x_{t,p,k}}{np \cdot n}$$

$$\bar{\bar{x}} = \frac{\sum_{t=1}^{nt} \sum_{p=1}^{np} \sum_{k=1}^n x_{t,p,k}}{nt \cdot np \cdot n}$$

Treating pH as a single treatment, regardless of temperature:

$$SS_{pH} = \sum_{t=1}^{nt} \sum_{p=1}^{np} \sum_{k=1}^n (\bar{x}_p - \bar{\bar{x}})^2$$

where
$$\bar{x}_p = \frac{\sum_{t=1}^{nt} \sum_{k=1}^n x_{t,p,k}}{nt \cdot n}$$

Considering interaction between pH and temperature as a treatment:

$$\begin{aligned} SS_{int} &= \sum_{t=1}^{nt} \sum_{p=1}^{np} \sum_{k=1}^n \left(Effect_{temp\#t \cdot pH\#p} \right)^2 \\ &= \sum_{t=1}^{nt} \sum_{p=1}^{np} \sum_{k=1}^n \left(\bar{x}_{t,p} - \left(\bar{x}_t + \bar{x}_p - \bar{\bar{x}} \right) \right)^2 \end{aligned}$$

(i) Now construct the analysis of variance table as below:

Source	df	Sum of Squares	Mean Square	F-Value	P(F>Fstat)
Treatment					
Temp					
pH					
Interaction					
Error					
Total					

(8)

- (ii) From the above table, comment on testing of the following 4 null hypotheses:
- No difference among treatments – Compare the result with part (a) of the question.
 - No differences among temperatures.
 - No differences among pH values.
 - No interactions among temperature and pH.

(2)

(c) Plot [mRNA] as a function of pH for each of the two cell types, with a given symbol representing a corresponding temperature. [mRNA] should be shown as average/mean with error bars showing standard deviations. Comment on the comparison of the results obtained in (b) with observations from the plots.

(2)