



IIT PALAKKAD

ഭാരതീയ സാങ്കേതികവിദ്യാ സ്ഥാപനം പാലക്കാട്  
ഭാരതീയ പ്രौദ്യോഗിക സംസ്ഥാന പാലക്കാട്  
Indian Institute of Technology Palakkad  
Nurturing Minds For a Better World

## DEPARTMENT OF DATA SCIENCE

### DS5610 BUSINESS ANALYTICS

#### TEST 1 (OPEN BOOK/NOTES)

21 February 2025

Marks: 30 (Weight: 15 Marks)

Duration: 50 Mins

#### PLEASE READ THE FOLLOWING CAREFULLY

- **Answer all the questions.** There are no choices.
- **Answer all parts of the question together.** For example, if Q. No. 2 contains sub-questions such as 2a, 2b and 2c, all these questions should be answered together in sequence in one place. **Please do not write 2a on page 1 and 2c on the last page. SUCH ANSWERS WILL NOT BE EVALUATED.**
- Draw diagrams neatly and legibly.
- **THIS IS AN OPEN-BOOK EXAMINATION.** Textbook, hand-written notes in bounded form, and printed slide handouts are allowed. **NO OTHER MATERIAL, LOOSE SHEETS** are allowed.
- The use of a scientific calculator is permitted. However, the use of mobile phone, laptop and other electronic device are **NOT PERMITTED** during the examination.
- No exchange of scientific calculators, books, printed slides, loose sheets, etc. are permitted.
- If you find any of the data missing for any of the questions, **MAKE YOUR OWN ASSUMPTIONS AND HIGHLIGHT THE SAME BY DRAWING A BOX AROUND IT.**
- **WRITE THE ANSWER IN A BLUE/BLACK PEN.** Writing by Pencil or Red/Green Pen is not allowed.

**Q. No. 1.** The “Carseats” dataset in the “ISLR” library records the sales of child car seats at 400 different stores. The dataset provides information on the following variables, as discussed in Table 1 below.

**Table 1:** Description of Variables

Sales	Unit sales (in thousands) at each location
CompPrice	Price charged by competitor at each location
Advertising	Local advertising budget for company at each location (in thousands of dollars)
Price	Price company charges for car seats at each site
ShelveLoc	A factor with levels “Bad”, “Medium” and “Good” indicating the quality of the shelving location

A linear regression model (Model 1) is built to predict sales based on the available information from the other variables. The model summary is presented below in Table 2. It is easy to understand that ShelveLocGood and ShelveLocMedium are dummy variables automatically created by R. Note that ShelveLocGood takes value 1 when the shelving location is good and 0, otherwise. Similarly, ShelveLocMedium takes value 1 when the shelving location is medium and 0, otherwise. The R-squared and Adjusted R-squared for the fitted Model 1 are 0.7782 and 0.7754, respectively.

**Table 2:** Summary of Regression Analysis based on Model 1

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	3.8444	0.5645	6.81	3.60e - 11
Price	-0.0936	0.0035	-26.74	< 2e - 16
ShelveLocGood	4.7293	0.1999	23.66	< 2e - 16
ShelveLocMedium	1.7728	0.1639	10.81	< 2e - 16
CompPrice	0.0937	0.0054	17.37	< 2e - 16
Advertising	0.1203	0.0101	11.88	< 2e - 16

Later, another linear regression model (Model 2) incorporating the interaction between Price and ShelveLoc is fitted on the same data set. The model summary is presented below in Table 3. The R-squared and Adjusted R-squared for the fitted Model 2 are 0.7790 and 0.7751, respectively.

**Table 3:** Summary of Regression Analysis based on Model 2

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	3.2262	0.8205	3.932	9.95e - 05
Price	-0.0889	0.0061	-14.627	< 2e - 16
ShelveLocGood	5.0876	0.9739	5.224	2.85e - 07
ShelveLocMedium	2.6940	0.8200	3.285	0.0011
CompPrice	0.0944	0.0054	17.339	< 2e - 16
Advertising	0.1207	0.0101	11.900	< 2e - 16
Price × ShelveLocGood	-0.0032	0.0082	-0.390	0.6971
Price × ShelveLocMedium	-0.0080	0.0070	-1.146	0.2525

Based on the available results, answer the following questions (No need to re-fit the models again in R):

- Which linear regression model is better? Why? [2]
- Based on your chosen model, interpret the regression coefficient estimates corresponding to Price and ShelveLoc in detail. [5]
- Are all the predictors relevant in this context? Discuss based on your chosen model. [2]
- Suppose you are provided with the following information (Table 4) from 3 new stores. Prepare a simple prediction dashboard with predicted sales for each of the 3 new stores obtained from your chosen model. How will you use such a prediction dashboard? Discuss briefly. [5]

**Table 4:** Details from New Stores

Store No	Price	ShelveLoc	CompPrice	Advertising
1	110	Bad	115	5
2	115	Medium	120	5
3	125	Good	120	15

- e. Now, the actual sales data is provided to you in Table 5. How will you assess the performance of your chosen model based on this data set? Discuss in detail. [4]

**Table 5:** Recorded Sales from 3 New Stores

Store No	1	2	3
Actual Sales	4.5	6.5	11.5

- f. The linear regression model uses here only a limited number of predictors. Identify at least 2 more predictors that can be useful in this context. [2]

## Q. No. 2

- a. For the “Car seats” case described above in Q. No. 1, explain precisely and briefly how the framework of Business Analytics can be applied. [3]
- b. Varun Beverages Limited (VBL) is the largest bottling company of PepsiCo’s beverages in the world outside the USA. An automatic filling machine fills the 7UP bottles (the lemon flavoured carbonated water) in one of their bottling plants. A random sample of 20 bottles results in a sample variance of fill volume of  $s^2 = 0.0153$  (ml) $^2$ . If the variance of fill volume exceeds 0.01 (ml) $^2$ , an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the VBL has a problem with underfilled or overfilled bottles? Use a significance level of 0.05, and assume that fill volume has a normal distribution.
- What is the parameter of interest for estimation in the above situation? [0.5]
  - Write the null and alternate hypothesis. [2]
  - What type of test would this be? Why? [1]
  - Identify the suitable test statistic for this situation [1]
  - The actual p-value turned out to be 0.0649. What is your inference? [1.5]
  - In this case, what would be the relation between computed test statistic and tabulated statistic value? [1]

\*\*\*GOOD LUCK\*\*\*