DATASET 1

Παράδειγμα 8.4 Ο υπεύθυνος εκπαίδευσης ενός ναυπηγείου θέλει να συγκρίνει την αποτελεσματικότητα δύο εναλλακτικών μεθόδων που χρησιμοποιούνται για την εκπαίδευση των ηλεκτροσυγκολητών της εταιρείας. Για το λόγο αυτό χωρίζει τυχαία 26 υποψήφιους για εκπαίδευση ηλεκτροσυγκολητές σε 2 ισοπληθείς ομάδες Α και Β για να εκπαιδευθούν με τις μεθόδους 1 και 2 αντίστοιχα. Κατά τη διάρκεια της εκπαίδευσης κάποιοι εκπαιδευόμενοι αποχώρησαν για λόγους ανεξάρτητους με το πρόγραμμα εκπαίδευσης και τελικά το πρόγραμμα ολοκλήρωσαν 12 άτομα της ομάδας Α και 10 άτομα της ομάδας Β. Μετά την ολοκλήρωση του προγράμματος οι εκπαιδευθέντες υποβλήθηκαν σε μιά κοινή για όλους πρακτική άσκηση προκειμένου να αξιολογηθούν οι γνώσεις και οι ικανότητες που απέκτησαν. Η βαθμολογία της απόδοσής τους στην πρακτική άσκηση δίνεται στον παρακάτω Πίνακα

ΒΑΘΜΟΛΟΓΙΑ ΕΚΠΑΙΔΕΥΘΕΝΤΩΝ

ΟΜΑΔΑ Α (x_i)	70	93	82	90	77	86	79	84	98	73	81	85
ΟΜΑΔΑ Β (y_i)	89	78	94	83	88	80	91	92	87	97		

Με βάση τα στοιχεία αυτά να συγκριθεί, σε επίπεδο σημαντικότητας 0.05, η αποτελεσματικότητα των δύο εναλλακτικών μεθόδων εκπαίδευσης. Με άλλα λόγια να εξετασθεί αν θα μπορούσε ο υπεύθυνος της εταιρείας να ισχυρισθεί, σε επίπεδο σημαντικότητας 0.05, ότι μετά την εκπαίδευση τα μέλη και των δύο ομάδων έχουν το ίδιο επίπεδο γνώσεων και ικανοτήτων.

Υποθέτουμε ότι η βαθμολογία ακολουθεί την κανονική κατανομή και οι διακυμάνσεις των δύο πληθυσμών μπορούν να υποτεθούν ίσες.

DATASET 2

11.25 In order to compare two computer software packages, a manager has 10 individuals use each software package to perform a standard set of tasks typical of those encountered in the office. Of course, in carrying out the comparison the manager was careful to use individuals who did not have an established preference or skill with either type of software, and five individuals were randomly selected to use Software A first while the other five used Software B first. The time required to perform the standard set of tasks, to the nearest minute, is reported in Table 11.3. Test the null hypothesis that there is no difference between the mean time required to perform the standard tasks by the two software packages, using the 5 percent level of significance.

Ans. Reject H_0.

Table 11.3 Time Required to Perform a Standard Set of Tasks Using Two Software Packages (Nearest Minute)

Individual	1	2	3	4	5	6	7	8	9	10
Software A	12	16	15	13	16	10	15	17	14	12
Software B	10	17	18	16	19	12	17	15	17	14

DATASET 3

Fifteen trainees in a technical program are randomly assigned to three different types of instructional approaches, all of which are concerned with developing a specified level of skill in computer-assisted design. The achievement test scores at the conclusion of the instructional unit are reported in Table 13.4, along with the mean performance score associated with each instructional approach. Use the analysis of variance procedure in Section 13.1 to test the null hypothesis that the three sample means were obtained from the same population, using the 5 percent level of significance for the test.

Table 13.4 Achievement Test Scores of Trainees under Three Methods of Instruction

Instructional method	Test scores					Total scores	Mean test scores
A_1	86	79	81	70	84	400	80
A_2	90	76	88	82	89	425	85
A_3	82	68	73	71	81	375	75

DATASET 4

THE RANDOMIZED BLOCK DESIGN (TWO-WAY ANALYSIS WITHOUT INTERACTION)

13.13 The designs produced by four automobile designers are evaluated by three product managers, as reported in Table 13.16. Test the null hypothesis that the average ratings of the designs do not differ, using the 1 percent level of significance.

Ans. Critical F ($df = 3, 6$) = 9.78. Computed $F = 12.29$. Therefore, reject the null hypothesis that $\alpha_k = 0$ for all treatment (column) effects.

Table 13.16 Ratings of Automobile Designs

Evaluator	Designer				Total (T_j)	Mean ($\bar{X}_j$)
	1	2	3	4		
A	87	79	83	92	341	85.25
B	83	73	85	89	330	82.50
C	91	85	90	92	358	89.50
Total (T_k)	261	237	258	273	Grand total $T = 1{,}029$	
Mean rating ($\bar{X}_k$)	87.0	79.0	86.0	91.0		Grand mean $\bar{X}_T = 85.75$

DATASET 5

LINEAR REGRESSION ANALYSIS

14.1 Suppose an analyst takes a random sample of 10 recent truck shipments made by a company and records the distance in miles and delivery time to the nearest half-day from the time that the shipment was made available for pick-up. Construct the scatter plot for the data in Table 14.1 and consider whether linear regression analysis appears appropriate.

Table 14.1 Sample Observations of Trucking Distance and Delivery Time for 10 Randomly Selected Shipments

Sampled shipment	1	2	3	4	5	6	7	8	9	10
Distance (X), miles	825	215	1,070	550	480	920	1,350	325	670	1,215
Delivery time (Y), days	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

DATASET 6

Table 15.3 reports single-family house prices, with a random sample of 10 houses taken from each of three housing subdivisions. As indicated, in addition to the subdivision and the price, the square footage of each house and each lot have been collected. With price being the dependent variable, carry out a backward stepwise regression analysis using computer software. Also obtain a residual plot for the final regression model and a correlation matrix of all the simple correlation coefficients as the basis for the solutions to the supplementary problems that follow. Use the binary coding scheme in Section 15.3 for the two indicator variables that are required for coding the housing subdivision.

Table 15.3 Single-Family House Prices in Three Subdivisions

Sampled house	Price	Living area sq. ft	Lot size, sq. ft	Subdivision
1	102,200	1,500	12,000	A
2	103,950	1,200	10,000	A
3	87,900	1,200	10,000	A
4	110,000	1,600	15,000	A
5	97,000	1,400	12,000	A
6	95,700	1.200	10,000	A
7	113,600	1,600	15,000	A
8	109,600	1,500	12,000	A
9	110,800	1,500	12,000	A
10	90,600	1,300	12,000	A
11	109,000	1,600	13,000	B
12	133,000	1,900	15,000	B
13	134,000	1,800	15,000	B
14	120,300	2,000	17,000	B
15	137,000	2,000	17,000	B
16	122,400	1,700	15,000	B
17	121,700	1,800	15,000	B
18	126,000	1,900	16,000	B
19	128,000	2,000	16,000	B
20	117,500	1,600	13,000	B
21	158,700	2,400	18,000	C
22	186,800	2,600	18,000	C
23	172,400	2,300	16,000	C
24	151,200	2,200	16,000	C
25	179,100	2,800	20,000	C
26	182,300	2,700	20,000	C
27	195,850	3,000	22,000	C
28	168,000	2,400	18,000	C
29	199,400	2,500	20,000	C
30	163,000	2,400	18,000	C

Obtain the multiple regression equation for estimating house price based on all three variables of house size, lot size, and location. Test the multiple regression model for significance at the 5 percent level.