11.7 The following is a portion of a classic data set called the "pilot plot data" in *Fitting Equations to Data* by Daniel and Wood, published in 1971. The response *y* is the acid content of material produced by titration, whereas the regressor *x* is the organic acid content produced by extraction and weighing.

y	x	y	x
76	123	70	109
62	55	37	48
66	100	82	138
58	75	88	164
88	159	43	28

(b) Fit a simple linear regression; estimate a slope and intercept.

b_0: *the slop* ; b_1: *the intercept*

$$b_1 = \frac{\sum x_i y_i - n\,\bar{X}\bar{Y}}{\sum x_i^2 - \bar{X}^2}, \qquad b_0 = \bar{Y} - b_1\bar{X}$$

$\bar{X} = \frac{123+\cdots+28}{10} = 99.9$, $\bar{Y} = \frac{76+\cdots+43}{10} = 67$.

$\sum x_i y_i = 74058$; $\sum x_i^2 = 119969$

Thus, $b_1 = \frac{74058-(10)(99.9)(67)}{119969-(10)(99.9)} = 0.3538$

$$b_0 = 67 - (0.3533)(99.91) = 31.71$$

The regression line:

$$\widehat{Y_i} = 31.71 + 0.3533\, x_i$$

11.9 A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.

Advertising Costs ($)	Sales ($)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

(b) Find the equation of the regression line to predict weekly sales from advertising expenditures.

(c) Estimate the weekly sales when advertising costs are 35$.

$\widehat{Y}_i = b_0 + b_1 x$, $b_1 = \dfrac{\sum x_i y_i - n\,\bar{X}\bar{Y}}{\sum x_i^2 - \bar{X}^2}$, $b_0 = \bar{Y} - b_1 \bar{X}$

$n = 12;\ \bar{X} = 34.1667$, $\bar{Y} = 453.75$.

$\sum x_i y_i = 191325;\quad \sum x_i^2 = 15650$

Thus, $b_1 = \frac{191325-(12)(34.1667)(453.75)}{15650-(12)(34.1667)^2} = 3.22$

$$b_0 = 453.75 - (3.22)(34.1667\ \ ) = 343.7$$

The equation of regression line is:
$\widehat{Y}_i = 343.7 + 3.22\ x_i$.

Point Estimation of weekly sales when advertising costs are $35
$x_k = 35,\ \widehat{Y}_k = 343.7 + 3.22\ (35)$= 456.434.

11.21 With reference to Exercise 11.9
a) Test the hypothesis that $\beta_1 = 5$ against the alternative that $\beta_1 < 5$. Use a 0.025 level of significance.
b) And construct a 99% confidence interval for β_1.
c) Find and interpret the Coefficient of Determination

x	y	$\hat{y}_i$	$y_i - \hat{y}_i$	$y_i - \bar{y}$
40	385	472.538	-87.5381	-68.75
20	400	408.122	-8.1218	-53.75
25	395	424.226	-29.2259	-58.75
20	365	408.122	-43.1218	-88.75
30	475	440.330	34.6701	21.25
50	440	504.746	-64.7462	-13.75
40	490	472.538	17.4619	36.25
20	420	408.122	11.8782	-33.75
50	560	504.746	55.2538	106.25
40	525	472.538	52.4619	71.25
25	480	424.226	55.7741	26.25
50	510	504.746	5.2538	56.25

a)

1. Hypothesis:

$$H_0: \beta_1 = 5\ \ v.s\ \ H_1: \beta_1 < 5$$

2. Test statistic:

$$T = \frac{b_1 - \beta_{10}}{\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} = \frac{3.22 - 5}{1.24} = -1.435$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 25226\ ;\ \hat{\sigma} = \sqrt{\frac{SSE}{(n-2)}} = \sqrt{\frac{25226}{10}} = 50.2255$$

$$S(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \frac{50.2255}{\sqrt{1640.6092}} = 1.24$$

3. Decision:
   Reject H_0 if $T < -t_{(\alpha, n-2)}$
   $T = -1.435 \nless -t_{(0.025,10)} = -2.228$, so we cannot reject H_0.

   b) **99% C.I for $\boldsymbol{\beta_1}$:** $b_1 - t_{(\frac{\alpha}{2}, n-2)} s(b_1) \leq \boldsymbol{\beta_1} \leq b_1 + t_{(\alpha/2, n-2)} s(b_1)$

   $b_1 = 3.22\ ;\ \propto = 0.01;\ \ s(b_1) = 1.24\ ;\ t_{0.005,10} = 3.169$

   with confidence coefficient 99 , we estimate $\boldsymbol{\beta_1} \in [\mathbf{-0.71, 7.15}]$

c) Coefficient of Determination

$$R^2 = 1 - \frac{SSE}{SST};\ \ SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

SST= 42256, $R^2 = 1 - \frac{25226}{42256} = 0.4030$

This means that the 40.30% of change in the mean sales for retail merchant is by advertising expenditures.

11.53 The following data represent the chemistry grades for a random sample of 12 freshmen at a certain college along with their scores on an intelligence test administered while they were still seniors in high school.

Student	Test Score, x	Chemistry Grade, y
1	65	85
2	50	74
3	55	76
4	65	90
5	55	85
6	70	87
7	65	94
8	70	98
9	55	81
10	70	91
11	50	76
12	55	74

(a) Compute and interpret the sample correlation coefficient.

a) $r = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} = \frac{\sum x_i y_i - n\bar{X}\bar{Y}}{\sqrt{(\sum x_i^2 - n\bar{X}^2)(\sum y_i^2 - n\bar{Y}^2)}}$

$\bar{X} = 60.4167, \bar{Y} = 84.25 \;\; ; \sum x_i^2 = 44475; \sum y_i^2 = 85905; \sum x_i y_i = 61685$

Thus, $r = \frac{61685 - (12)(60.4167)(84.25)}{\sqrt{(44475 - (12)(60.4167))(85905 - (12)(84.25)^2)}} = 0.862$

Strong positive relationship (Strong positive correlation).

One-Way Analysis of Variance

13.1 Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. The following are the tensile-strength measurements in kilograms per square centimeter $x10^{-1}$:

Machine					
1	**2**	**3**	**4**	**5**	**6**
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1

Perform the analysis of variance at the 0.05 level of significance and indicate whether or not the mean tensile strengths differ significantly for the six machines.

$\propto = 0.05, k = 6, n = 4$ " 6 rubber machines"

1. $H_0: \mu_1 = \mu_2 = \cdots = \mu_6.\;\; V.S \;\;\; H_1: At\ least\ one\ \mu_i\ is\ different.$
2. $Test\ Statistic$:

i	1	2	3	4	5	6
Total $y_{i.}$	68.8	68.7	72.7	71	73.7	72.1
Mean $\bar{y}_{i.}$	17.2	17.175	18.175	17.75	18.425	18.025

$$\bar{y}_{..} = \frac{68.8 + \cdots + 72.1}{24} = 17.79$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = 4\sum_{i=1}^{6}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSA = 4[(17.2 - 17.79)^2 + \cdots + (18.025 - 17.79)^2] = 5.338$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2,$$

$$SST = (17.5 - 17.79)^2 + \cdots + (20.1 - 17.79)^2 = 67.978$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2, SSE = SST - SSA$$

$$SSE = 67.978 - 5.338 = 62.64$$

$$MSE = \frac{SSE}{K(n-1)} = \frac{62.64}{6(3)} = 3.48$$

ANOVA Table

source	**df**	**SS**	**MS**	**F**
Treatments	k-1=5	SSA=5.338	$MSA = S_1^2 = 1.068$	$\frac{1.068}{3.48} = 0.307$
Error	K(n-1)=18	SSE=62.64	$MSE = S^2 = 3.48$	
Total	Kn-1=23	SST=67.978		

3. Rejection Region (R.R):

$$Reject\ H_0\ if\ F > f_{\propto,k-1,k(n-1)} = f_{0.05,5,18} = 2.773$$

4. Decision:

Since $F = 0.307 \ngtr 2.773$ $we\ cannot\ reject\ H_0$.

$thus, the\ mean\ tensile\ strengths\ is\ the\ same\ for\ the\ 6\ machines.$

H.W 13.3 In an article "Shelf-Space Strategy in Retailing," published in *Proceedings: Southern Marketing Association*, the effect of shelf height on the supermarket sales of canned dog food is investigated. An experiment was conducted at a small supermarket for a period of 8 days on the sales of a single brand of dog food, referred to as Arf dog food, involving three levels of shelf height:

knee level, waist level, and eye level. During each day, the shelf height of the canned dog food was randomly changed on three different occasions. The remaining sections of the gondola that housed the given brand were filled with a mixture of dog food brands that were both familiar and unfamiliar to customers in this particular geographic area. Sales, in hundreds of dollars, of Arf dog food per day for the three shelf heights are given. Based on the data, is there a significant difference in the average daily sales of this dog food based on shelf height? Use a 0.01 level of significance.

Shelf Height		
Knee Level	Waist Level	Eye Level
77	88	85
82	94	85
86	93	87
78	90	81
81	91	80
86	94	79
77	90	87
81	87	93

$$\propto = 0.01\ , k = 3\ , n = 8$$

1. $H_0: \mu_1 = \mu_2 = \mu_3.$ $V.S$ $H_1: At\ least\ one\ \mu_i\ is\ different.$
2. $Test\ Statistic$

i	Knee	Waist	Eye
Total $y_{i.}$	648	727	677
Mean $\bar{y}_{i.}$	81	90.875	84.625

$$\bar{y}_{..} = 85.5$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = 8\sum_{i=1}^{3}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSA = 8[(81 - 85.5)^2 + (90.875 - 85.5)^2 + (84.625 - 85.5)^2] = 399.25$$

$$MSA = \frac{SSA}{(k-1)} = \frac{399.25}{(3)} = 199.62$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$SSE = (77 - 81)^2 + \cdots + (81 - 81)^2 + (88 - 90.875)^2 + \cdots + (87 - 90.875)^2$$
$$+(85 - 84.625)^2 + \cdots + (93 - 84.625)^2 = 288.8$$

$$SST = SSE - SSA$$

$$SST = 288.8 + 5.338 = 199.62$$

$MSE = \frac{SSE}{K(n-1)} = \frac{288.8}{3(7)}$ =13.75

Thus, $F = \frac{MSA}{MSE} = \frac{199.62}{13.75} = 14.52$

ANOVA Table

source	**df**	**SS**	**MS**	**F**
Treatments	k-1=2	SSA=399.2	199.62	14.52
Error	K(n-1)=21	SSE=288.8	13.75	
Total	Kn-1=23	SST=688		

1. Rejection Region (R.R):

$$Reject\ H_0\ if\ F > f_{\propto,k-1,k(n-1)} = f_{0.01,2,21} = 5.78$$

2. Decision:

Since $F = 14.52 > 5.78$ $we\ reject\ H_0$.
thus, there is a significant difference in the average daily sales of this dog food based on shelf height.

Two-Factor Analysis of Variance

14.1 An experiment was conducted to study the effects of temperature and type of oven on the life of a particular component. Four types of ovens and 3 temperature levels were used in the experiment. Twenty-four pieces were assigned randomly, two to each combination of treatments, and the following results recorded.

	Oven			
Temperature (°F)	O_1	O_2	O_3	O_4
500	227	214	225	260
	221	259	236	229
550	187	181	232	246
	208	179	198	273
600	174	198	178	206
	202	194	213	219

Using a 0.05 level of significance, test the hypothesis that:
(a) different temperatures have no effect on the life of the component;
(b) different ovens have no effect on the life of the component;
(c) the type of oven and temperature do not interact.

$$a = 3\,, b = 4\,, n = 2\,, \propto = 0.05$$

$\bar{\boldsymbol{y}}_{\boldsymbol{ij.}}$: $\bar{y}_{11.} = 224$ $\bar{y}_{12.} = 236.6$ $\bar{y}_{13.} = 230.5$ $\bar{y}_{14.} = 244.5$

$\bar{y}_{21.} = 197.5$ $\bar{y}_{22.} = 180$ $\bar{y}_{23.} = 215$ $\bar{y}_{24.} = 215$

$\bar{y}_{31.} = 188$ $\bar{y}_{32.} = 196$ $\bar{y}_{33.} = 195.5$ $\bar{y}_{34.} = 212.5$

$\bar{\boldsymbol{y}}_{\boldsymbol{i..}}$: $\bar{y}_{1..} = 233.875$ $\bar{y}_{2..} = 213$ $\bar{y}_{3..} = 198.$

$\bar{\boldsymbol{y}}_{\boldsymbol{.j.}}$: $\bar{y}_{.1.} = 203.1667\,, \bar{y}_{.2.} = 204.11, \bar{y}_{.3.} = 213.667\,, \bar{y}_{.4.} = 238.833$

$\bar{\boldsymbol{y}}_{...} = 214.958$

c)

1. $\boldsymbol{H_0}$: *There is no interaction bettween the diffferent tempreatures*

$and\ the\ different\ types\ of\ ovens.$

$\boldsymbol{H_1}: There\ is\ interaction\ bettween\ the\ diffferent\ tempreatures$
$and\ the\ different\ types\ of\ ovens.$

$2. Test\ Statistic:$

$$SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{\boldsymbol{y}}_{ij.} - \bar{\boldsymbol{y}}_{i..} - \bar{\boldsymbol{y}}_{.j.} + \bar{\boldsymbol{y}}_{...}\right)^2$$

$$= 2\,[(224 - 233.875\ - 203.1667\ +)^2 + \cdots$$
$$+\ (212.5 - 198\ - 238.833\ + 214.958)^2] = 3126$$

$$MSAB = \frac{3126}{6} = 521$$

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\boldsymbol{y}_{ijk} - \bar{\boldsymbol{y}}_{ij.}\right)^2 = (227 - 224)^2 + \cdots + (219 - 212.5)^2 = 3833$$

$$MSE = \frac{SSE}{ab(n-1)} = \frac{3833}{12} = 319.5$$

$$F_3 = \frac{MSAB}{MSE} = \frac{521}{319.5} = 1.63$$

4.Decision:

$F_3 = 1.63\ \ngtr 2.996 = f_{0.05,6,12}$

We cannot reject $\boldsymbol{H_0}$ i.e ther is no interaction.

a) **1.** $\boldsymbol{H_0}: There\ is\ no\ difference\ in\ the\ mean\ of\ the\ comonent\ life$
$of\ the\ 3\ tempreatures\ .$

$\boldsymbol{H_1}: At\ least\ one\ mean\ different.$

2.Test Statistic:

$$SSA = b\,n\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2$$

$$= 8[(233.875 - 214.958)^2 + \cdots + (198 - 214.958)^2] = 5194.$$

$$MSA = \frac{519}{2} = 2597$$

$$F = \frac{MSA}{MSE} = \frac{2597}{312.5} = 8.128$$

3.Desision:

Since $F = 8.128 > 3.9 = f_{0.05,2,12}$ we reject $\boldsymbol{H_0}$.

There is a significant difference in the mean of component life of the 3 temperatures.

b) **1.** $\boldsymbol{H_0}$: $There\ is\ no\ difference\ in\ the\ mean\ of\ the\ comonent\ life$
$of\ the\ 3\ tempreatures\ .$

$\boldsymbol{H_1}$: $At\ least\ one\ mean\ different.$

2.Test Statistic:

$$SSB = a\,n \sum_{i=1}^{b} \left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2$$

$$= (203.1667 - 214.958)^2 + \cdots + (238.833 - 214.958)^2 = 4963$$

$$MSB = \frac{MSB}{MSE} = \frac{1654.4}{319.5} = 5.18$$

3.Decision:

Since $F = 5.18 > 3.49 = f_{0.05,3,12}$ we reject $\boldsymbol{H_0}$.

There is a significant difference in the mean of component life of the 4 ovens types.

H.W 14.3 Three strains of rats were studied under 2 environmental conditions for their performance in a maze test. The error scores for the 48 rats were recorded.

	Strain					
Environment	**Bright**		**Mixed**		**Dull**	
Free	28	12	33	83	101	94
	22	23	36	14	33	56
	25	10	41	76	122	83
	36	86	22	58	35	23
Restricted	72	32	60	89	136	120
	48	93	35	126	38	153
	25	31	83	110	64	128
	91	19	99	118	87	140

Use a 0.01 level of significance to test the hypothesis that:
(a) there is no difference in error scores for different environments;
(b) there is no difference in error scores for different strains;
(c) the environments and strains of rats do not interact.

c) **1.** $\boldsymbol{H_0}$: $There\ is\ no\ interaction\ bettween\ the\ diffferent\ environments$
$and\ diffferent\ \ strains\ of\ rats\ .$

$\boldsymbol{H_1}$: $There\ is\ interaction\ bettween\ the\ diffferent\ the\ diffferent\ environments$
$and\ diffferent\ \ strains\ of\ rats\ .$

2.Test Statistic:

$$MSAB = 617.6, \qquad MSE = 1004.6, \qquad F_3 = 0.61$$

3.Decision:

$F_3 = 0.61 \ \ngtr f_{0.01,2,42}$

We cannot reject $\boldsymbol{H_0}$ i.e ther is no interaction.

a) **1.** $\boldsymbol{H_0}$: $There\ is\ no\ difference\ in\ the\ mean\ of\ error\ scores\ for\ different\ environments$

$\boldsymbol{H_1}$: $At\ least\ one\ mean\ different.$

2.Test Statistic:

$$SSA = b\,n \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$MSA = 14875.5$$

$$F = \frac{MSA}{MSE} = 14.81$$

3.Desision:

Since $F = 14.81 > f_{0.01,1,42}$ we reject $\boldsymbol{H_0}$.

There is a significant difference in the error scores mean for different environments.

b) **1.** $\boldsymbol{H_0}$: $There\ is\ no\ difference\ in\ the\ mean\ of\ error\ scores\ for\ different\ strains\ of\ rats.$

$\boldsymbol{H_1}$: $At\ least\ one\ mean\ different.$

2.Test Statistic:

$$SSB = a\,n \sum_{i=1}^{b} \left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2$$

$$MSB = \frac{MSB}{MSE} = 9077.1$$

$$F_2 = 9.04$$

3.Decision:

Since $F_2 = 9.04 > f_{0.01,2,42}$ we reject $\boldsymbol{H_0}$.

Thus, there is a significant difference in the error scores mean for different strains of rats.