12.1.{2, 3, 9}, 12.2.{16, 17, 22a, 24}, 4.6.{88, 89, 94}, 12.5.{58, 59}

12.1.2 The article **"Exhaust Emissions from Four-Stroke Lawn Mower Engines"** (*J. of the Air and Water Mgmnt. Assoc.*, **1997: 945–952)** reported data from a study in which both a baseline gasoline mixture and a reformulated gasoline were used. Consider the following observations on age (yr) and NO_x emissions (g/kWh):

Engine	1	2	3	4	5	6	7	8	9	10
Age	0	0	2	11	7	16	9	0	12	4
Baseline	1.72	4.38	4.06	1.26	5.31	.57	3.37	3.44	.74	1.24
Reformulated	1.88	5.93	5.54	2.67	6.53	.74	4.94	4.89	.69	1.42

Construct scatterplots of No_x emmisions versus age. What appears to be the nature of the relationship between these two variables? [*Note*: The authors of the cited article commented on the relationship.]

12.1.3 Bivariate data often arises from the use of two different techniques to measure the same quantity. As an example, the accompanying observations on x = hydrogen concentration (ppm) using a gas chromatography method and y = concentration using a new sensor method were read from a graph in the article **"A New Method to Measure the Diffusible Hydrogen Content in Steel Weldments Using a Polymer Electrolyte-Based Hydrogen Sensor"** (*Welding Res.*, July 1997: 251s–256s).

x	47	62	65	70	70	78	95	100	114	118
y	38	62	53	67	84	79	93	106	117	116
x	124	127	140	140	140	150	152	164	198	221
y	127	114	134	139	142	170	149	154	200	215

Construct a scatterplot. Does there appear to be a very strong relationship between the two types of concentration measurements? Do the two methods appear to be measuring roughly the same quantity? Explain your reasoning.

■

12.1.9 The flow rate y (m^3/min) in a device used for air-quality measurement depends on the pressure drop x (in. of water) across the device's filter. Suppose that for x values between 5 and 20, the two variables are related according to the simple linear regression model with true regression line $y = -.12 + .095x$.

(a) What is the expected change in flow rate associated with a 1-in. increase in pressure drop? Explain.

(b) What change in flow rate can be expected when pressure drop decreases by 5 in.?

(c) What is the expected flow rate for a pressure drop of 10 in.? A drop of 15 in.?

(d) Suppose $\sigma = .025$ and consider a pressure drop of 10 in. What is the probability that the observed value of flow rate will exceed .835? That observed flow rate will exceed .840?

(e) What is the probability that an observation on flow rate when pressure drop is 10 in. will exceed an observation on flow rate made when pressure drop is 11 in.?

■

12.2.16 The article **"Characterization of Highway Runoff in Austin, Texas, Area"** (*J. of Envir. Engr.*, **1998: 131–137)** gave a scatterplot, along with the least squares line, of x = rainfull volume (m^3) and y = runoff volume (m^3) for a particular location. The accompanying values were read from the plot.

x	5	12	14	17	23	30	40	47	55	67	72	81	96	112	127
y	4	10	13	15	15	25	27	46	38	46	53	70	82	99	100

(a) Does a scatterplot of the data support the use of the simple linear regression model?

(b) Calculate point estimates of the slope and intercept of the population regression line.

(c) Calculate a point estimate of the true average runoff volume when rainfall volume is 50.

(d) Calculate a point estimate of the standard deviation σ.

(e) What proportion of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall?

12.2.17 No-fines concrete, made from a uniformly graded coarse aggregate and a cement-water paste, is beneficial in areas prone to excessive rainfall because of its excellent drainage properties. The article **"Pavement Thickness Design for No-Fines Concrete Parking Lots," *J. of Trans. Engr.*, 1995: 476–484)** employed a least squares analysis in studying how y = porosity (%) is related to x = unit weight (pcf) in concrete specimens. Consider the following representative data:

x	99.0	101.1	102.6	103.0	105.4	107.0	108.7	110.8
y	28.8	27.9	27.0	25.2	22.8	21.5	20.9	19.6
x	112.1	112.4	113.6	113.8	115.1	115.4	120.0	
y	17.1	18.9	16.0	16.7	13.0	13.6	10.8	

Relevant summary quantities are $\Sigma x_i = 1640.1$, $\Sigma y_i = 299.8$, $\Sigma x_i^2 = 179,849.73$, $\Sigma x_i y_i = 32,308.59$, $\Sigma y_i^2 = 6430.06$.

(a) Obtain the equation of the estimated regression line. Then create a scatterplot of the data and graph the estimated line. Does it appear that the model relationship will explain a great deal of the observed variation in y?

(b) Interpret the slope of the least squares line.

(c) What happens if the estimated line is used to predict porosity when unit weight is 135? Why is this not a good idea?

(d) Calculate the residuals corresponding to the first two observations.

(e) Calculate and interpret a point estimate of σ.

(f) What proportion of observed variation in porosity can be attributed to the approximate linear relationship between unit weight and porosity?

Solve by hand.

12.2.22a Calcium phosphate cement is gaining increasing attention for use in bone repair applications. The article **"Short-Fibre Reinforcement of Calcium Phosphate Bone Cement"** (*J. of Engr. in Med.*, **2007: 203–211**) reported on a study in which polypropylene fibers were used in an attempt to improve fracture behavior. The following data on x = fiber weight (%) and y = compressive strength (MPa) was provided by the article's authors.

x	0.00	0.00	0.00	0.00	0.00	1.25	1.25	1.25	1.25
y	9.94	11.67	11.00	13.44	9.20	9.92	9.79	10.99	11.32
x	2.50	2.50	2.50	2.50	2.50	5.00	5.00	5.00	5.00
y	12.29	8.69	9.91	10.45	10.25	7.89	7.61	8.07	9.04
x	7.50	7.50	7.50	7.50	10.00	10.00	10.00	10.00	
y	6.63	6.43	7.03	7.63	7.35	6.94	7.02	7.67	

(a) Fit the simple linear regression model to this data. Then determine the proportion of observed variation in strength that can be attributed to the model relationship between strength and fiber weight. Finally, obtain a point estimate of the standard deviation of ϵ, the random deviation in the model equation.

■

12.2.24 The invasive diatom species *Didymosphenia geminata* has the potential to inflict substantial ecological and economic damage in rivers. The article **"Substrate Characteristics Affect Colonization by the Bloom-Forming Diatom** *Didymosphenia geminata* **(***Aquatic Ecology***, 2010: 33–40)** described an investigation of colonization behavior. One aspect of particular interest was whether y = colony density was related to x = rock surface area. The article contained a scatterplot and summary of a regression analysis. Here is representative data:

x	50	71	55	50	33	58	79	26	69	44	37	70	20	45	49
y	152	1929	48	22	2	5	35	7	269	38	171	13	43	185	25

(a) Fit the simple linear regression model to this data, predict colony density when surface area = 70 and when surface area = 71, and calculate the corresponding residuals. How do they compare?

(b) Calculate and interpret the coefficient of determination.

(c) The second observation has a very extreme y value (in the full data set consisting of 72 observations, there were two of these). This observation may have had a substantial impact on the fit of the model and subsequent conclusions. Eliminate it and recalculate the equation of the estimated regression line. Does it appear to differ substantially from the equation before the deletion? What is the impact on r^2 and s?

4.6.88 A sample of 15 female collegiate golfers was selected and the clubhead velocity (km/hr) while swinging a driver was determined for each one, resulting in the following data **("Hip Rotational Velocities During the Full Golf Swing," *J. of Sports Science and Medicine*, 2009: 296–299)**:

$$
\begin{array}{ccccc}
69.0 & 69.7 & 72.7 & 80.3 & 81.0 \\
85.0 & 86.0 & 86.3 & 86.7 & 87.7 \\
89.3 & 90.7 & 91.0 & 92.5 & 93.0
\end{array}
$$

The corresponding z percentiles are

$$
\begin{array}{ccccc}
-1.83 & -1.28 & -0.97 & -0.73 & -0.52 \\
-0.34 & -0.17 & 0.0 & 0.17 & 0.34 \\
0.52 & 0.73 & 0.97 & 1.28 & 1.83
\end{array}
$$

Construct a normal probability plot and a dotplot. Is it plausible that the population distribution is normal?

4.6.89 The accompanying sample consisting of $n = 20$ observations on dielectric breakdown voltage of a piece of epoxy resin appeared in the article **"Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution (***IEEE Trans. on Dielectrics and Elec. Insul.,*** 1996: 43–55)**. The values of $(i - .5)/n$ for which z percentiles are needed are $(1 - .5)/20 = .025$, $(2 - .5)/20 = .075$, ..., and $.975$. Would you feel comfortable estimating population mean voltage using a method that assumed a normal population distribution?

Observation	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
z percentile	-1.96	-1.44	- 1.15	-.93	-.76	-.60	-.45	-.32	-.19	-.06
Observation	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
z percentile	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

∎

4.6.94 The accompanying observations are precipitation values during March over a 30-year period in Minneapolis-St. Paul.

.77	1.20	3.00	1.62	2.81	2.48	1.74	.47	3.09	1.31
1.87	.96	.81	1.43	1.51	.32	1.18	1.89	1.20	3.37
2.10	.59	1.35	.90	1.95	2.20	.52	.81	4.75	2.05

(a) Construct and interpret a normal probability plot for this data set.

(b) Calculate the square root of each value and then construct a normal probability plot based on this transformed data. Does it seem plausible that the square root of precipitation is normally distributed?

(c) Repeat part (b) after transforming by cube roots.

12.5.58 The Turbine Oil Oxidation Test (TOST) and the Rotating Bomb Oxidation Test (RBOT) are two different procedures for evaluating the oxidation stability of steam turbine oils. The article **"Dependence of Oxidation Stability of Steam Turbine Oil on Base Oil Composition"** (*J. of the Society of Tribologists and Lubrication Engrs.,* **Oct. 1997: 19–24**) reported the accompanying observations on $x =$ TOST time (hr) and $y =$ RBOT time (min) for 12 oil specimens.

TOST	4200	3600	3750	3675	4050	2770	4870	4500	3450	2700	3750	3300
RBOT	370	340	375	310	350	200	400	375	285	225	345	285

(a) Calculate and interpret the value of the sample correlation coefficient (as do the article's authors).

(b) How would the value of r be affected if we had let $x =$ RBOT time and $y =$ TOST time?

(c) How would the value of r be affected if RBOT time were expressed in hours?

(d) Construct normal probability plots and comment.

(e) Carry out a test of hypotheses to decide whether RBOT time and TOST time are linearly related.

∎

12.5.59 Toughness and fibrousness of asparagus are major determinants of quality. This was the focus of a study reported in **"Post-Harvest Glyphosate Application Reduces Toughening, Fiber Content, and Lignification of Stored Asparagus Spears"** (*J. of the Amer. Soc. of Hort. Science*, **1988: 569–572**). The article reported the accompanying data (read from a graph) on $x = $ shear force (kg) and $y = $ percent fiber dry weight.

x	46	48	55	57	60	72	81	85	94
y	2.18	2.10	2.13	2.28	2.34	2.53	2.28	2.62	2.63
x	109	121	132	137	148	149	184	185	187
y	2.50	2.66	2.79	2.80	3.01	2.98	3.34	3.49	3.26

$n = 18, \Sigma x_i = 1950, \Sigma x_i^2 = 251,970, \Sigma y_i = 47.92, \Sigma y_i^2 = 130.6074, \Sigma x_i y_i = 5530.92.$

(a) Calculate the value of the sample correlation coefficient. Based on this value, how would you describe the nature of the relationship between the two variables?

(b) If a first specimen has a larger value of shear force than does a second specimen, what tends to be true of percent dry fiber weight for the two specimens?

(c) If shear force is expressed in pounds, what happens to the value of r? Why?

(d) If the simple linear regression model were fit to this data, what proportion of observed variation in percent fiber dry weight could be explained by the model relationship?

(e) Carry out a test at significance level .01 to decide whether there is a positive linear association between the two variables.

Solve by hand.