Appendix A: Solution Key

This appendix provides brief solutions to some of the problems in the book. Often only the final answer or value is provided without any explanation or justification. Given the sometimes subjective nature of regression analysis, other equally valid answers can also exist. As well, although the author has strived to verify that the answers are correct, it is inevitable that one or two stray errors may appear. This should be borne in mind when comparing answers. If you believe that there is indeed an error in the solution key, please let the author know so that appropriate corrections can be made.

Chapter 1

(1) F; (2) F; (3) T; (4) F; (5) T; (6) F; (7) T; (8) F; (9) T; (10) T; (11) F; (12) T; (13) F; (14) F; (15) T; (16) T; (17) F; (18) T; (19) T; (20) T

(21) (a) $\mu = 4.3$, mode $= 5$, median $= 5$; (b) $\sigma^2 = 5.12$, MAD $= 1.84$, range $= 7$; (c) Q1 $= 2$, Q2 $= 5$, Q3 $= 5$

(24) Maybe!

Chapter 2

(1) T; (2) F; (3) T; (4) T; (5) T; (6) F; (7) T; (8) T; (9) F; (10) F; (11) T; (12) T; (13) F; (14) T; (15) T; (16) T; (17) F

(18) (a) $\mathbb{S} = \{$Q♥, Q♠, Q♣$\}$; $\mathbb{F} = \{\{\}, \{$Q♥$\}, \{$Q♠$\}, \{$Q♣$\}, \{$Q♥, Q♠$\}, \{$Q♥, Q♣$\}, \{$Q♠, Q♣$\}, \{$Q♥, Q♠, Q♣$\}\}$; $P(\{\}) = 0$; $P(\{$Q♥$\}) = \frac{1}{3}$, $P(\{$Q♠$\}) = \frac{1}{3}$; $P(\{$Q♣$\}) = \frac{1}{3}$; $P(\{$Q♥, Q♠$\}) = \frac{2}{3}$; $P(\{$Q♥, Q♣$\}) = \frac{2}{3}$; $P(\{$Q♠, Q♣$\}) = \frac{2}{3}$; $P(\{$Q♥, Q♠, Q♣$\}) = 1$; (b) $\frac{1}{3}$; (c) $\mu = 0$; $\sigma^2 = 2$.

Y.A.W. Shardt, *Statistics for Chemical and Process Engineers*,
DOI 10.1007/978-3-319-21509-9

Table A.1 Answers for question 27 in Chap. 2

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$
Bias	0	0	$0.5\,\mu$	$^1/_3\mu$	0
σ^2	σ^2	$0.5\sigma^2$	$1.25\sigma^2$	$0.\overline{2}\sigma^2$	σ^2 / N
MSE	σ^2	$0.5\sigma^2$	$1.25\sigma^2 + 0.25\,\mu$	$0.\overline{2}\sigma^2 + 0.\overline{1}\mu$	σ^2 / N

(20) (a) $\mathbb{S} = \{\text{HH, HT, TH, TT}\}$; $\mathbb{F} = \{\{\}, \{\text{HH}\}, \{\text{HT}\}, \{\text{TH}\}, \{\text{TT}\}, \{\text{HH, HT}\}, \{\text{HH, TH}\}, \{\text{HH, TT}\}, \{\text{HT, TH}\}, \{\text{HT, TT}\}, \{\text{TH, TT}\}, \{\text{HH, HT, TH}\}, \{\text{HH, HT, TT}\}, \{\text{HH, TH, TT}\}, \{\text{HT, TH, TT}\}, \{\text{HH, HT, TH, TT}\}\}$; $\text{P}(\{\}) = 0$, $\text{P}(\{\text{HH}\}) = 0.3$, $\text{P}(\{\text{HT}\}) = 0.3$, $\text{P}(\{\text{TH}\}) = 0.2$, $\text{P}(\{\text{TT}\}) = 0.2$, $\text{P}(\{\text{HH, HT}\}) = 0.6$, $\text{P}(\{\text{HH, TH}\}) = \text{P}(\{\text{HH, TT}\}) = \text{P}(\{\text{HT, TH}\}) = \text{P}(\{\text{HT, TT}\}) = 0.5$, $\text{P}(\{\text{TH, TT}\}) = 0.4$, $\text{P}(\{\text{HH, HT, TH}\}) = \text{P}(\{\text{HH, HT, TT}\}) = 0.8$, $\text{P}(\{\text{HH, TH, TT}\}) = \text{P}(\{\text{HT, TH, TT}\}) = 0.7$, $\text{P}(\{\text{HH, HT, TH, TT}\}) = 1$.

(21) (a) No; (b) No; (c) Yes, $\mu = 2.5$, $\sigma^2 = 25/12$, $\text{E}(|x|) = 2.5$.

(23) $E(2X - 4Y) = -6$; $E(3XY) = 6$; $E(X^2) = 3$.

(27) See Table A.1.

(29) (a) Yes; (b) No; (c) Yes; (d) No.

(34) (b) $\mu = 0.5\pi$; (c) $4.84 \leq \psi \leq 6.41$, yes; (d) sampled mean is not equal to the true value; (e) $P(0.25 < X < 0.5) = 0.046$.

Chapter 3

(1) F; (2) T; (3) F; (4) T; (5) F; (6) T; (7) T; (8) F; (9) F; (10) T; (11) F; (12) F; (13) T; (14) T; (15) F; (16) T; (17) F; (18) T; (19) T; (20) T

(23) The solution is given as

$$\vec{y} = \langle y_3, y_4, \ldots, y_{100} \rangle^T$$

$$\mathcal{A} = \begin{bmatrix} -y_2 & -y_1 & u_2 & u_1 \\ -y_3 & -y_2 & u_3 & u_2 \\ \vdots & \vdots & \vdots & \vdots \\ -y_{99} & -y_{98} & u_{99} & u_{98} \end{bmatrix}$$

$$\vec{\beta} = \langle \alpha_1, \alpha_2, \beta_1, \beta_2 \rangle^T$$

(24) (c) $\hat{R} = 2.15\ \Omega$, $\hat{\sigma} = 0.85$; (d) $2.1 \pm 0.2\ \Omega$; (e) Yes; (f) Fit is not good, as the errors are increasing with current.

(25) (b) $\hat{R} = 2.044\ \Omega$; (c) An examination of the residuals (especially as a function of the current, I) and how close to normality they are; (d) Instrument error is often proportional to the magnitude of the measured value.

(27) (a + b) $\beta_0 = 276 \pm 6.8$; $\beta_1 = -1.98 \pm 1.0$; $\beta_2 = 0.012 \pm 0.041$; $\beta_3 = 0.00018 \pm 0.00045$; (c) 229.5 ± 3.3, 229.5 ± 8.5; (d) Yes; (e) $R^2 = 0.963$, F-score $= 87.23$, No.

(29) (b + c) $\ln \hat{K} = 5.3 \pm 0.45$, $\hat{b} = 0.54 \pm 0.28$; (d) $128 \leq K \leq 317$; $\hat{b}$ stays the same; (e) $\hat{\sigma} = 0.103$, $R^2 = 0.8789$; (g) Yes; (h) $R = 0.467\ \mathrm{kg \cdot m^{-0.5} \cdot s^{-1}}$; (i) $\hat{K} = 181 \pm 67, \hat{b} = 0.46 \pm 0.25; R^2 = 0.8905, \hat{\sigma} = 7.397$; (j) Yes, $R = 0.518\ \mathrm{kg \cdot m^{-0.5} \cdot s^{-1}}$.

(31) (a) No, plot the errors for the two runs using different symbols; (c) Yes.

Chapter 4

(1) T; (2) T; (3) F; (4) F; (5) F; (6) F; (7) F; (8) F; (9) T; (10) T; (11) T; (12) T; (13) T; (14) F; (15) T; (16) T; (17) F; (18) F; (19) F; (20) T; (21) (a) No; (b) Yes. (23) (a) I = ACD = CDF = AF; (b) (two examples) A = F = CD = ACDF, B = ABF = ABCD = ABCDF; (c) III; (d) (one of many) I = ABCE = ACDF = BDEF.

(25) (a brief outline of the solution) A fractional factorial design with centre point replicates. Blocking and randomisation should also be considered.

(26) (a) $I = x_1 x_2^2 x_3 = x_1^2 x_2 x_3^2$; (b) $x_1 = x_1^2 x_2^2 x_3 = x_2 x_3^2$, $x_2 = x_1 x_3 = x_1^2 x_2^2 x_3^2$, $x_3 = x_1 x_2^2 x_3^2 = x_1^2 x_2$, $x_1^2 = x_2^2 x_3 = x_1 x_2 x_3^2$, $x_2^2 = x_1 x_2 x_3 = x_1^2 x_3^2$, $x_3^2 = x_1 x_2^2 = x_1^2 x_2 x_3$, $x_1 x_2 = x_1^2 x_3 = x_2^2 x_3^2$, $x_2 x_3 = x_1 x_3^2 = x_1^2 x_2^2$, $x_1 x_2^2 x_3 = x_1^2 x_2 x_3^2 = I$; (c) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3$

(29) (b) E = −ABCD; c) V; d) BD, AE, DE, BE, and E.

(31) (a) $y = 100 - 4x_1^2 - 12x_2^2 - 9x_3^2$; (b) 100.

Chapter 5

(1) T; (2) F; (3) T; (4) T; (5) F; (6) F; (7) F; (8) T; (9) T; (10) F; (11) T; (12) T; (13) T; (14) F; (15) T; (16) F; (17) T; (18) F; (19) T; (20) F; (24) AR(1). (25) seasonal MA(1) with $s = 3$ and a normal AR(4).

Chapter 6

(1) F; (2) T; (3) T; (4) F; (5) F; (6) t; (7) T; (8) T; (9) T; (10) T; (11) F; (12) F; (13) T; (14) F; (15) T; (16) F; (17) T; (18) F; (19) F; (20) T.

(24) left graph: 10, right graph: 5. (25) Model is not adequate. (26) Model is adequate.

References

Ashley R (1988) On the relative worth of recent macroeconomic forecasts. Int J Forecast 4:363–376

Barnett V, Lewis T (1994) Outliers in statistical data, 3rd edn. Wiley, Chichester

Bloomfield P (2000) Fourier analysis of time series: an introduction, 2nd edn. Wiley, New York

Box GE, Jenkins GM (1970) Time series analysis, forecasting, and control. Holden-Day, Oakland

Box GE, Hunter WG, Hunter JS (1978) Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley, New York

Brillo J (2007) Excel for scientists and engineers: numerical methods. Wiley, Hoboken

Chambers JM, Cleveland WS, Kleiner B, Tukey P (1983) Graphical methods for data analysis. Wadsworth International Group, Belmont

Daniel C, Wood FS (1980) Fitting equations to data, 2nd edn. Wiley, New York

Davies L, Gather U (1993) The identification of multiple outliers. J Am Stat Assoc 88 (423):782–792

Dean JA (1999) Lange's handbook of chemistry, 15th edn. McGraw-Hill, Inc., New York

Elliott JA, Prickett RC, Elmoazzen HY, Porter KR, McGann LE (2007) A multisolute osmotic virial equation for solutions of interest in biology. J Phys Chem B 111:1775–1785

Franke J, Härdle WK, Hafner CM (2011) Statistics of financial markets: an introduction, 3rd edn. Springer, Heidelberg. doi:10.1007/978-3-642-16521-4

Gerhart PM, Gross RJ, Hochstein JI (1992) Fundamentals of fluid mechanics. Addison-Wesley Publication Co., Reading

Hald A (2003) A history of probability and statistics and their application before 1750. Wiley, Hoboken

Hannan EJ, Deistler M (2012) The statistical theory of linear systems. Society of Industrial and Applied Mathematics, Philadelphia (Originally published: Wiley, New York, 1988)

Hare LB (1988) In the soup: a case study to identify contributors to filling variability. J Qual Technol 20:36–43

Harvey AC (1991) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge

Harvey G (2013) Excel® 2013 All-in-One for dummies. Wiley, Hoboken

Hassler U (1994) The sample autocorrelation function of I(1) processes. Statistical Papers—Statistische Hefte 35:1–16

Hawkins DM (1980) Identification of outliers. Chapman and Hall, London

Hinkelmann K, Kempthorne O (2007) Design and analysis of experiments, vol I, II, III. Wiley, Hoboken

Y.A.W. Shardt, *Statistics for Chemical and Process Engineers*,
DOI 10.1007/978-3-319-21509-9

Hodge VJ, Austin J (2004) A survey of outlier detection methodologies. Artif Intell Rev 22:85–126
Huang B, Kadali R (2008) Dynamic modeling, predictive control, and performance monitoring. Springer, London
Hunt BR, Lipsman RL, Rosenberg J (2014) A guide to MATLAB: for beginners and experienced users: updated for MATLAB 8 and Simulink 8, 3rd edn. Cambridge University Press, Cambridge, UK
Hyndman RJ, Fan Y (1996) Sample quantiles in statistical packages. Am Stat 50(4):361–365
Jochem M, Körber C (1987) Extended phase diagrams for the ternary solutions $H_2O - NaCl - glycerol$ and $H_2O - NaCl - hydroxyethylstarch$ (HES) determined by DSC. Cryobiology 24:513–536
Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis, 6th edn. Prentice Hall, Upper Saddle River
Kallenberg O (2002) Foundations of modern probability. Springer, New York
Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45
Kalman RE, Bucy RS (1961) New results in filtering and prediction theory. Trans ASME J Basic Eng 83:95–108
Khintchine A (1934) Korrelationstheorie der stationären stochastischen prozesse (correlation theory of stocastic processes). Math Ann 109(1):604–615. doi:10.1007/BF01449156
Lin B, Recke B, Knudsen JK, Jørgensen SB (2007) A systematic approach for soft sensor development. Comput Chem Eng 31:419–425
Ljung L (1999) System identification theory for the user. Prentice Hall, Inc., Upper Saddle River
Montgomery DC (1991) Design and analysis of experiments, 3rd edn. Wiley, New York
Montgomery DC, Peck EA (1982) Introduction to linear regression analysis, 1st edn. Wiley, New York
Montgomery DC, Runger GC (2007) Applied statistics and probability for engineers, 4th edn. Wiley, Hoboken
Montgomery DC, Jennings CL, Kulahci M (2008) Introduction to time series analysis and forecasting. Wiley, Hoboken
Myers RH (1971) Response surface methodology. Allyn and Bacon, Inc., Boston
Nelson CR (1972) The prediction performance of the FRB-MIT-PENN model of the U.S. Economy. Am Econ Rev 62(5):902–917
Ogunnaike BA (2010) Random phenomena: fundamentals of probability and statistics for engineers. CRC Press, Boca Raton
Prickett RC, Elliott JA, McGann LE (2010) Application of the osmotic virial equation in cyrobiology. Cryobiology 2010:30–42
Prickett RC, Elliott JA, McGann LE (2011) Application of the multisolute osmotic virial equation to solutions containing electrolytes. J Phys Chem B 115:14531–14543
Priestley MB (1981) Spectral analysis and time series: Vol. 1: univariate series and vol 2: multivariate series, prediction, and control. Academic, New York
Ross-Rodriguez LU (2009) Cellular osmotic properties and cellular responses to cooling. University of Alberta, Edmonton
Schlittgen R (2012) Einführung in die statistik analyse und modellierung von daten, 12th edn. Oldenbourg Wissenschaftsverlag GmbH, Munich
Schmuller J (2013) Statistical analysis with Excel® for dummies, 3rd edn. Wiley, Hoboken
Seber GA, Wild CJ (1989) Nonlinear regression. Wiley, New York
Shardt Y (2012a) Data quality assessment for closed-loop system identification and forecasting with application to soft sensors. Doctoral thesis, University of Alberta, Department of Chemical and Materials Engineering, Edmonton, Alberta, Canada. doi: http://hdl.handle.net/10402/era.29018

Shardt YA (2012b) Data quality assessment for closed-loop system identification and forecasting with application to soft sensors. University of Alberta, Edmonton. doi: http://hdl.handle.net/10402/era.29018

Shardt YA, Huang B (2011) Closed-loop identification with routine operating data: effect of time delay and sampling time. J Process Control 21:997–1010

Shardt YA, Huang B (2014) Minimal required excitation for closed-loop identification: implications for PID control loops. In: ADCONIP conference proceedings, Hiroshima, Japan, pp 296–301. doi:http://www.nt.ntnu.no/users/skoge/prost/proceedings/adconip-2014/pdf/SUBS61TO80/0069/0069_FI.pdf

Sheynin O (2004) History of the theory of probability to the beginning of the 20th century. NG Verlag, Berlin

Shumway RH, Stoffer DS (2011) Time series analysis and its applications with R examples, 3rd edn. Springer, New York. doi:10.1007/978-1-4419-7865-3

Sizemore J, Mueller JP (2015) MATLAB for dummies. Wiley, Hoboken

Söderström T, Gustavsson I, Ljung L (1975) Identifiability conditions for linear systems operating in closed loop. Int J Control 21(2):243–255

Sokal RR, Rohlf FJ (1995) Biometry: the principles and practice of statistics in biological research, 3rd edn. W. H. Freeman and Company, New York

Steel RG, Torrie JH (1980) Principles and procedures of statistics: a biometrical approach, 2nd edn. McGraw-Hill Kogakusha, Ltd., Tokyo

Stoica P, Moses R (2005) Spectral analysis of signals. Prentice Hall, Upper Saddle River

Tufte ER (1997) Visual and statistical thinking: displays of evidence for making decisions. Graphics Press LLC., Cheshire

Tufte ER (2001) The visual display of quantitative information. Graphics Press LLC., Cheshire

Varberg DE (1963) The development of modern statistics. The Mathematics Teacher 56(4): 252–257

Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified Periodograms. IEEE Trans Audio Elecotroacustics AU-15(2):70–73

Wichern DW (1973) The behaviour of the sample autocorrelation function for an integrated moving average process. Biometrika 60(2):235–239. doi:http://www.jstor.org/stable/2334535?origin=JSTOR-pdf

Zhu Y (2001) Multivariable system identification for process control. Elsevier Science Ltd., Oxford

Zorn ME, Gibbons RD, Sonzogni WC (1997) Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. Anal Chem 69(15):3069–3075

Subject Index

An index of Excel and Matlab topics is listed as a separate index located after this one on p. 413.

Y.A.W. Shardt, *Statistics for Chemical and Process Engineers*,
DOI 10.1007/978-3-319-21509-9

Index of Excel and MATLAB Topics

This index only considers Excel and Matlab functions that are used in this book. The main index is found on p. 407.

Y.A.W. Shardt, *Statistics for Chemical and Process Engineers*,
DOI 10.1007/978-3-319-21509-9