
Engineering Statistics

A Handbook of Statistical Inference, Regression, and Experimentation for Engineers

Mansur M. Arief
Industrial & Systems Engineering
King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia

Preface

This book is an accompanying material for the course ISE 315, Engineering Statistics, offered by the Industrial & Systems Engineering Department at King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia. It has grown into a handbook of statistical inference, regression, and experimentation for engineers, organized in six parts of content plus a worksheets part. Part I builds the foundations: the role of statistics in engineering, how engineering data are collected and summarized, point estimation, and sampling distributions with the central limit theorem. Part II develops inference for a single sample: confidence intervals for means, variances, and proportions, and the full machinery of hypothesis testing. Part III compares two populations: two-sample procedures for means (with variances known, pooled, unequal, and paired), proportions, and variances. Part IV builds regression models: simple linear regression with its inference and diagnostics, then multiple linear regression. Part V treats the design and analysis of experiments: single-factor analysis of variance with multiple comparisons and blocking, and two-factor and factorial experiments. Part VI closes with statistics in practice: a software-and-tools chapter that maps every method in the book to its spreadsheet and Python implementation. A final part collects the printable worksheets, the formula sheet, and the statistical tables used throughout the course.

The reference textbook for the course is Montgomery and Runger, *Applied Statistics and Probability for Engineers*, 6th edition (Wiley). This book follows the notation of that textbook exactly (μ , σ^2 , $\bar{X}$, S^2 , $\hat{p}$, $z_{\alpha/2}$, $t_{\alpha/2, n-1}$, and so on), so the two can be read side by side, and a margin reference at the start of each section points to the matching textbook section. The book is nevertheless written so that it can be read on its own. You do not need the lecture recordings or the textbook next to you while you read it, although both are useful. Every method

is explained from the beginning, every formula is derived so that you can see where it comes from, and every kind of problem that appears on the homework, the quizzes, and the major exams is solved here in full, step by step.

How this book is organized

Each chapter follows the order of the course lectures. The body of the page carries the explanation, the derivations, and the worked examples. The wide margin carries supporting material, so you always know where to look:

Definitions. Every technical term, when it is first used, is defined in the margin in green.

Figures. Small statistical plots—dot diagrams, distribution curves, table excerpts with the used cell highlighted—sit in the margin, next to the sentence that refers to them, so the picture and the words stay together. Larger panels that you must study closely are placed in the body at full width.

References. Pointers to the matching section of Montgomery and Runger, and to other sources, are placed in the margin at the start of each section.

Lecture pointers. This book is handed out chapter by chapter as the course proceeds, so each chapter tells you exactly which lecture covers what. A green margin note (“Lecture *N*”) marks the first section of each lecture’s coverage, and a green rule across the body with the words “Lecture *N* starts here” marks the exact point where that lecture’s material begins. If you read the block between two rules before the corresponding class, you arrive prepared.

Startup lens. Gold margin remarks that reread the section’s method from the point of view of someone building a new venture: designing an A/B test, reading churn data, deciding when a pilot result is strong enough to act on.

Engineering judgment. Maroon margin remarks that show where your engineering knowledge—process physics, measurement limits, capacity constraints, failure modes—enters the statistical analysis as an input, not an afterthought.

Safety lens. Maroon margin remarks that connect the method to AI safety testing and to verification and validation practice: how a confidence interval bounds a failure rate, why a test plan for an autonomous system is a sample-size problem, what a p-value does and does not certify.

Numbered procedures

New in this book, every multi-step recipe in the course is set out as a numbered, boxed *Procedure*: constructing a confidence interval, the seven steps of a hypothesis test, fitting a regression line, building an ANOVA table. Procedures are numbered within their chapter (“Procedure 3.2” and so on), and the lecture slides use the *same* numbers, so when a slide says “apply Procedure 3.2,” this book is where that procedure lives, with its derivation and worked examples around it. When you solve a problem, citing the procedure you used is part of showing your work.

Examples, figures, and tables

Every worked *example* is numbered within its chapter (“Example 2.3” and so on) so it can be referenced in lecture and in your own notes. Inside each example a small *Solution.* rule separates the problem statement from the worked-out solution, leaving you a natural place to attempt the problem yourself before reading on. Every figure and every table is also numbered and listed in the front matter, so the *List of Figures* and the *List of Tables* can be used as a quick visual index of the book. Equations that you will reuse are numbered as well, and the lecture slides cite those equation numbers, so the book and the slides always point to the same formula.

Practice: quizzes, activities, and exercises

Four kinds of practice are built into every chapter:

Pre-class quizzes. short concept checks placed at the start of the chapter, before the first lecture it covers. The questions appear consecutively in a small block; the answers are collected in a shaded box at the bottom of the same block, so you can answer all of them first and then check.

In-class activities. each lecture has a matching in-class activity (ICA), printed at the end of its chapter as a standalone page that can be extracted and used directly in class. ICAs are 15–20 minutes of team work, mostly reasoning rather than computing, with space to write and an exit check at the end.

End-of-chapter exercises. full computational problems in the style of the homework and the quizzes, arranged from basic descriptive calculations up to complete inference problems. Every solution is worked in full, but is collected in the back of the book (the “Solutions to End-of-Chapter Exercises” chapter), so that the temptation to glance at the answer while attempting the problem is removed.

Exam-style problems. each chapter’s exercises include problems written in the exact format and difficulty of the major exams and the final, with point values marked, so the step from practice to exam holds no surprises.

Each chapter’s exercises close with a group called *assessing outside recommendations*: problems in which an AI assistant, a vendor, or a consultant has already produced a statistical analysis—and that analysis is wrong or incomplete in some way. Your task is to find the error, explain it, and correct it. Checking someone else’s statistics is a different skill from producing your own, and in practice it is the one you will use most.

Multiple-choice questions, where they appear, list one option per indented line below the stem, so the choices are easy to scan.

A note on numbers and examples

Money amounts, where they appear, are in Saudi riyals. The examples, quizzes, activities, and exercises are built around three recurring settings, chosen so that the same statistical method is seen doing three different jobs: *AI safety testing* (validating a perception model’s failure rate, red-team defect discovery, guardrail false-positive rates, autonomous-vehicle disengagement data), *supply chains and transportation* (port container dwell times, truck fleet fuel use, warehouse picking times, delivery lateness), and *building your own startup* (A/B tests on signup conversion, user churn, pilot-customer response times, pricing experiments). Small differences between your answers and the book’s usually come from rounding, or from the number of decimal places carried in the statistical tables, and are not errors.

This resource was prepared with the help of generative AI tools and reviewed and corrected by the instructor.

Contents

PART I FOUNDATIONS OF STATISTICAL INFERENCE

1	<i>Statistics in Engineering</i>	3
1.1	The Role of Statistics in Engineering	4
1.2	Populations, Samples, and Two Directions of Reasoning	5
1.3	How Engineers Collect Data	8
1.4	Summarizing Data with Numbers	10
1.5	Summarizing Data with Pictures	14
1.6	The Road Ahead: What This Course Covers	17
1.7	Statistical Thinking, Step by Step	18
	ICA 1.1: First Data, First Decisions	24
2	<i>Point Estimation and Sampling Distributions</i>	27
2.1	The Statistical Inference Framework	28
2.2	Point Estimators and Point Estimates	30
2.3	Sampling Distributions	31
2.4	Unbiasedness and the Standard Error of $\bar{X}$	32
2.5	Mean Squared Error: Combining Bias and Variance	34
2.6	The Central Limit Theorem	37
2.7	Practice with Estimators	41
2.8	The Sample Proportion	45
2.9	Estimators for Differences of Two Means	47
2.10	Differences of Two Proportions	50
	ICA 2.1: One Fleet, Many Averages	57
	ICA 2.2: Which Estimator Would You Ship?	59
	ICA 2.3: Two Carriers, One Contract	61

PART II INFERENCE FOR A SINGLE SAMPLE

3	<i>Confidence Intervals: Inference for a Single Sample</i>	65
3.1	From Point Estimate to Interval Estimate	66
3.2	Confidence Interval on the Mean, σ Known	67
3.3	What “95% Confident” Actually Means	69
3.4	One-Sided Confidence Bounds	71
3.5	Precision and the Choice of Sample Size	72
3.6	Reading Critical Values from the z Table	74
3.7	The t Distribution and the CI on the Mean, σ Unknown	74
3.8	Confidence Interval on the Variance	77
3.9	Large-Sample Confidence Interval on a Proportion	80
3.10	Prediction Interval for a Single Future Observation	83
3.11	Tolerance Intervals	84
3.12	Choosing the Right Interval	86
3.13	Worked Practice	87
	ICA 3.1: How Wide Is the Truth?	95
	ICA 3.2: Choose Your Distribution	97
	ICA 3.3: The Dashboard Audit	99
4	<i>Tests of Hypotheses: One Sample</i>	101
4.1	From Intervals to Tests: The Toolkit So Far	102
4.2	Statistical Hypotheses	103
4.3	Decision Errors: Type I and Type II	105
4.4	Test Statistics and the Rejection Region	107
4.5	P-Values	109
4.6	The Seven-Step Testing Procedure	112
4.7	Tests on the Mean of a Normal Distribution, Variance Known	113
4.8	Tests on the Mean of a Normal Distribution, Variance Unknown	114
4.9	Tests on the Variance of a Normal Distribution	116
4.10	Tests on a Population Proportion	118
4.11	The Duality Between Tests and Confidence Intervals	119
4.12	Type II Error, Power, and Sample Size	120
4.13	Choosing the Right Test	123
4.14	Practice: Three Problems, One Procedure	124
	ICA 4.1: Whose Burden of Proof?	131
	ICA 4.2: Run the Seven Steps	133
	ICA 4.3: Which Test, and How Sure Are We?	135

PART III COMPARING TWO POPULATIONS

5	<i>Inference for Two Samples</i>	139
5.1	From One Sample to Two	140
5.2	Two Means, Variances Known: the Two-Sample z	141
5.3	Two Means, Variances Unknown but Equal: the Pooled t	144
5.4	Two Means, Unequal Variances: the Welch t	146
5.5	Paired Samples: the Paired t	148
5.6	Two Proportions	152
5.7	Two Variances: the F Test	153
5.8	Which Procedure? A Decision Map	157
5.9	Major Exam 1 Review: Chapters 1–5 in One Map	158
5.10	Major Exam 1 Practice: Worked Problems	160
	ICA 5.1: Two Gates, One Bottleneck	170
	ICA 5.2: Paired or Not?	172
	ICA 5.3: The Test-Picker Tournament	174
	ICA 5.4: Mini Major	176

PART IV REGRESSION MODELING

6	<i>Simple Linear Regression</i>	181
6.1	Empirical Models and the Scatter Diagram	182
6.2	The Simple Linear Regression Model	183
6.3	The Method of Least Squares	185
6.4	Estimating the Error Variance	189
6.5	Properties of the Least Squares Estimators	190
6.6	Fitting the Line in Practice	192
6.7	Testing for Significance of Regression	195
6.8	The Analysis of Variance for Regression	197
6.9	R^2 and the Sample Correlation Coefficient	199
6.10	Confidence Intervals for the Slope and Intercept	201
6.11	Mean Response and Prediction of New Observations	202
6.12	Residual Analysis: Checking the Model	206
6.13	Regression on Transformed Variables	209
6.14	Abuses of Regression	211
	ICA 6.1: Draw the Line Before You Trust It	218
	ICA 6.2: One Line Through the Port	220
	ICA 6.3: Is the Trend Real or Noise?	222
	ICA 6.4: Two Intervals, One Decision	224

7	<i>Multiple Linear Regression</i>	227
7.1	From One Regressor to Several	228
7.2	The Matrix Form and Least Squares	230
7.3	Estimating σ^2 and the Standard Errors	234
7.4	ANOVA for MLR and the Overall F Test	236
7.5	R^2 and the Adjusted R^2	238
7.6	Reading a Regression Summary from Software	240
7.7	Tests and Intervals for Individual Coefficients	242
7.8	The Partial F Test: Extra Sum of Squares	245
7.9	Intervals for the Mean Response and for a New Observation	247
7.10	Multicollinearity and the Variance Inflation Factor	248
7.11	Indicator Variables	250
7.12	Interaction and Polynomial Terms	251
7.13	Building a Model Step by Step	252
7.14	Residual Diagnostics for MLR	254
	ICA 7.1: Two Regressors, One Port	260
	ICA 7.2: Which Knob Moves Churn?	262

PART V DESIGN AND ANALYSIS OF EXPERIMENTS

8	<i>Single-Factor Experiments: Analysis of Variance</i>	267
8.1	Why Not Repeated Pairwise Tests?	268
8.2	Designed Experiments and the Completely Randomized Design	270
8.3	The Effects Model	271
8.4	Partitioning Variability	272
8.5	The F Test and the ANOVA Table	274
8.6	Estimating Treatment Effects and Confidence Intervals	277
8.7	Multiple Comparisons After the F Test	279
8.8	Checking the Model: Residual Analysis	284
8.9	Choosing the Sample Size	285
8.10	The Random-Effects Model	287
8.11	The Randomized Complete Block Design	290
	ICA 8.1: Three Red Teams, One Verdict	304
	ICA 8.2: Which Flow Do We Ship?	306
	ICA 8.3: Block the Weekday	308

9	<i>Two-Factor and Factorial Experiments</i>	311
9.1	Why Factorial Experiments?	312
9.2	Main Effects and Interaction	314
9.3	The Two-Factor Effects Model	316
9.4	Partitioning the Variability	317
9.5	The Two-Factor ANOVA Table and Its F Tests	318
9.6	A Complete Hand Computation	321
9.7	When the Interaction Is Significant	323
9.8	Residual Analysis and Model Adequacy	326
9.9	One Observation per Cell	329
9.10	General Factorial Experiments	330
9.11	Two-Level Designs: The 2^k Workhorse	331
9.12	Single Replicates, Center Points, Blocking, and Fractions	336
9.13	Factorial Thinking in Engineering Practice	339
	ICA 9.1: Whose Model Is Better? It Depends	346
	ICA 9.2: Sixteen Cohorts to a Better Signup Page	348

PART VI STATISTICS IN PRACTICE

10	<i>Software and Tools</i>	353
10.1	From Tables to Tools	354
10.2	The Division of Labor	355
10.3	Confidence Intervals in Software	357
10.4	Hypothesis Tests in Software	361
10.5	Two-Sample Methods in Software	363
10.6	Regression in Software	366
10.7	ANOVA and Designed Experiments in Software	370
10.8	The Complete Method-to-Tool Map	372
10.9	Rounding: Tables versus Software	373
10.10	Using AI Assistants Responsibly	374
	ICA 10.1: Reading the Printout	381

PART VII WORKSHEETS AND APPENDICES

11	<i>Worksheets and Practice Problems</i>	385
	Worksheet 11.1: Homework 1	387
	Worksheet 11.2: Homework 2	394
	Worksheet 11.3: Homework 3	398
	Worksheet 11.4: Homework 4	403
	Worksheet 11.5: Homework 5	409
	Worksheet 11.6: Homework 6	416
	Worksheet 11.7: Homework 7	424
	Worksheet 11.8: Homework 8	431
A	<i>Formula Sheet</i>	441
B	<i>Statistical Tables</i>	457
	<i>Solutions to End-of-Chapter Exercises</i>	465

List of Figures

- 1.1 Two packing lines with the same average output but very different variability. The average alone cannot tell them apart. 4
- 1.2 The two directions of statistical reasoning. Probability deduces sample behavior from a known population; inference reasons back from an observed sample to the unknown population. 6
- 1.3 Dot diagram of the eight miss distances of Table 1.3. The stack at 0.6 marks the repeated value. 14
- 1.4 Two common histogram shapes. In the right-skewed shape the long tail pulls the mean above the median. 14
- 1.5 Histogram of the 40 container dwell times of Table 1.4. The longer right tail is typical of waiting-time data. 15
- 1.6 Comparative box plots of time to first booking under the two onboarding flows of Table 1.5. 16
- 1.7 The route through this course: from describing a sample to estimating parameters, bounding them, testing claims about them, and modeling and designing with them. The software chapter supports every stage. 18

- 2.1 The statistical inference framework. Probability reasons from a known population to a sample; inference reasons from an observed sample back to the unknown population. 29
- 2.2 Population distribution versus sampling distribution for $X \sim \text{Uniform}(0, 1)$. The population density (left) is flat and never changes. The sampling distribution of $\bar{X}$ (right) is centered at the same $\mu = 0.5$ but is bell-shaped, with spread σ^2/n that shrinks as n increases. 32

- 2.3 The standard error of $\bar{X}$ for $\sigma = 12$. Doubling precision requires quadrupling the sample size. 33
- 2.4 Two estimators of the same parameter $\theta = 10$. Estimator A (solid) is unbiased but spread out. Estimator B (dashed) is biased low by 0.4 but tightly concentrated, and its mean squared error is less than half of A 's. 35
- 2.5 Sampling distributions of three estimators of the mean engagement time when $\mu = 10$, $\sigma = 10$, $n = 25$ (Example 2.2). The sample mean is both centered on μ and concentrated; the single observation is centered but wide; the shrinkage estimator is concentrated but centered at $0.7\mu = 7$. 37
- 2.6 The Central Limit Theorem acting on a strongly right-skewed population (exponential with $\mu = 1$; dashed gray line at μ). Solid green: the exact sampling distribution of $\bar{X}$ for $n = 1, 5, 30$. Dashed maroon: the normal approximation $N(\mu, \sigma^2/n)$ from the CLT. By $n = 30$ the two curves nearly coincide. 39
- 2.7 The shaded area $P(-1.50 \leq Z \leq 1.50) = 0.8664$ for Example 2.4. 42
- 2.8 MSE of the two estimators in Example 2.5 as a function of the unknown μ . The curves cross at $|\mu| = 4.359$ cm. 44
- 2.9 Sampling distribution of $\hat{P}$ in Example 2.7, with the event $|\hat{P} - p| \leq 0.05$ shaded. 47
- 2.10 The deviation of $\bar{X}_1 - \bar{X}_2$ from the true difference in Example 2.8, with the event $|\text{deviation}| < 1.5$ shaded. 50
- 3.1 The standard normal density with the central area $1 - \alpha$ (green) between $-z_{\alpha/2}$ and $z_{\alpha/2}$, and area $\alpha/2$ in each tail (maroon). For a 95% interval, $\alpha = 0.05$ and $z_{0.025} = 1.96$. 68
- 3.2 Twenty 95% confidence intervals, each built from an independent sample of $n = 25$ from a population with $\mu = 50$ and $\sigma = 10$. Eighteen intervals (green) contain μ ; samples 11 and 17 (maroon) happened to produce unusually high and unusually low means, and their intervals miss. In the long run about one interval in twenty misses. 70
- 3.3 A one-sided bound puts all of α in a single tail, so the critical value is z_α , not $z_{\alpha/2}$. At 95%: $z_{0.05} = 1.645 < 1.96 = z_{0.025}$. 71
- 3.4 The margin of error $E = 1.96\sigma/\sqrt{n}$ at 95% confidence. Halving E (from 0.784σ to 0.392σ to 0.196σ) requires quadrupling n (from about 6 to 25 to 100): precision is bought at a square-root rate. 73

- 3.5 The t density for $\nu = 3$ and $\nu = 10$ degrees of freedom against the standard normal. Smaller ν means heavier tails, larger critical values, and wider confidence intervals; as $\nu \rightarrow \infty$ the t curve becomes the normal curve. 75
- 3.6 The chi-square density with $\nu = 9$ degrees of freedom. The distribution is right-skewed, so the two critical values ($\chi_{0.975,9}^2 = 2.700$ and $\chi_{0.025,9}^2 = 19.023$ for $\alpha = 0.05$) are not symmetric around the center, and the resulting CI for σ^2 is not symmetric around s^2 . 78
- 3.7 The variance factor $p(1 - p)$ peaks at $p = 0.5$ with value 0.25. Planning with $\hat{p} = 0.5$ in (3.8) is therefore the worst case: it can never under-deliver on precision. 82
- 4.1 Type I and type II errors for the dwell time test ($n = 64, \sigma = 16, \alpha = 0.05$). The maroon tails under the H_0 curve have total area α ; the gold area under the true-mean curve is β , computed in Example 4.7. 107
- 4.2 Rejection regions for the z test at significance level α . The shaded area is α in every panel; only its placement changes with the direction of H_1 . 108
- 4.3 Two-sided p -value: the total shaded area $2[1 - \Phi(|z_0|)]$, drawn for the dwell time statistic $z_0 = 2.50$. 110
- 4.4 Lower-tailed chi-square test with $\nu = 19$: the shaded rejection region lies below $\chi_{0.95,19}^2 = 10.117$; the observed $\chi_0^2 = 10.64$ falls just outside it. 118
- 4.5 β for Example 4.7(a): the probability that $\bar{X}$, distributed around the true mean 119, still lands inside the fail-to-reject interval (111.08, 118.92). 122
- 4.6 Operating characteristic curves for the two-sided dwell-time z test ($\mu_0 = 115, \sigma = 16, \alpha = 0.05$), computed from (4.10). Larger samples detect the same shift with much smaller β ; at $n = 168$ the shift to 119 is missed only about 10% of the time. 122
- 5.1 The two-sample framework. Two populations are observed through two independent random samples, and inference targets the *difference* or *ratio* of their parameters. 141
- 5.2 Two-sided rejection region at $\alpha = 0.05$ for Example 5.1. The observed $z_0 = -1.30$ does not reach either shaded tail. 142
- 5.3 Welch degrees of freedom (5.7) for $n_1 = n_2 = 10$ as the variance ratio grows. At equal variances $\nu = 18$; strongly unequal variances push ν toward $n - 1 = 9$. 147

- 5.4 The paired fuel-efficiency data of Example 5.4. Each line is one vehicle. The vehicles are spread over almost 4 km/L, yet every line tilts up by a consistent 0.4–0.9 km/L: pairing isolates exactly that tilt, while an independent-samples analysis must fight the full vertical spread. 151
- 5.5 The $F_{15,20}$ density with the upper $\alpha/2 = 0.025$ rejection region shaded, for Example 5.6; the observed $f_0 = 2.33$ falls just short of the critical value. The dashed $F_{5,5}$ curve shows how strongly the shape depends on the degrees of freedom. 156
- 5.6 Decision map for comparing two means. Ask the questions in order: pairing first, then knowledge of the variances, then their equality. Proportions and variances have their own procedures ((5.10) and (5.13)). 158
- 6.1 Three scatter-diagram patterns. Simple linear regression is appropriate for pattern (a). Pattern (b) calls for a transformation (Section 6.13); in pattern (c), x has little predictive value for y . 183
- 6.2 The simple linear regression model (6.2). At each value of x , the response Y has a normal distribution (gold curves, drawn sideways) centered on the population regression line (green) with the *same* variance σ^2 everywhere. 184
- 6.3 The verification-testing data of Example 6.1 with the fitted least squares line. The line passes exactly through $(\bar{x}, \bar{y}) = (27.5, 31.625)$, as every least squares line must. 188
- 6.4 The fuel data of Example 6.3 with the fitted line. The gold segments show two residuals: least squares chooses the line that minimizes the sum of the squares of all ten such vertical gaps. 194
- 6.5 If $\beta_1 = 0$, knowing x tells you nothing about Y . The test asks whether the observed slope is too far from zero to be noise. 195
- 6.6 Two data sets with the same kind of fitted line. Left: the points hug the line and the model explains most of the variability. Right: the trend is present but weak, and most of the variability remains in the residuals. 200
- 6.7 Confidence band for the mean response (inner, green) and prediction band for a new observation (outer, gold) for the ad-spend regression. Both bands are narrowest at $\bar{x} = 6$ and widen toward the edges of the data; the prediction band is wider everywhere because of the extra “1+” in (6.27). 204

- 6.8 Four residual patterns. (a) is what an adequate model produces. (b) violates constant variance: consider transforming y (Section 6.13). (c) means the true relationship is not a straight line: add curvature or transform. (d) shows an outlier: check the data before anything else. 207
- 6.10 Inside the data (solid) the line is supported by evidence. Beyond it, the fitted line (dashed) and the truth (maroon) are free to part ways, and nothing in the data can warn you. 211
- 6.9 The backlog data of Example 6.10. Left: in the original scale the growth is exponential and a straight line would fit badly. Right: after the transformation $y' = \ln y$, the trend is linear and ordinary least squares applies. 211
- 7.1 The fitted MLR surface for $k = 2$ regressors is a plane, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Each residual is the vertical distance from a data point (dots) to the plane. Least squares chooses the plane that minimizes the sum of the squared vertical distances. 230
- 7.2 As regressors are added one at a time, R^2 never falls, but R^2_{adj} peaks (here at $k = 4$) and then declines: the later regressors do not earn the degrees of freedom they cost. 239
- 7.3 With two regressors, $R_1^2 = r^2$, so $\text{VIF} = 1/(1 - r^2)$. The inflation is mild until $|r| \approx 0.8$ and then rises steeply: $|r| = 0.95$ already gives $\text{VIF} = 10.3$. 249
- 7.4 The cost of multicollinearity. Panel (a): two regressors that track each other almost perfectly. Panel (b): the estimate $\hat{\beta}_1$ (true value 2, dashed) refitted on twelve independent samples of the same size. With an orthogonal design the estimates cluster tightly; with the collinear design of panel (a) they swing between -1.4 and 5.3 , sometimes even changing sign. 250
- 7.5 One indicator variable, two parallel lines. The gap between the lines is the indicator coefficient, constant across all prices. 251
- 7.6 An interaction between distance and a route-type indicator: the lines have different slopes, so each extra kilometer costs more time in the city than on the highway. 252
- 7.7 A quadratic model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$ fitted by ordinary linear least squares: the model is linear in the β 's even though the curve is not a line. 252

- 7.8 Reading residual-vs-fitted plots in MLR. (a) A random horizontal band: the model is adequate. (b) A funnel: the error variance is not constant; transform the response. (c) A systematic curve: the mean model is missing a quadratic or interaction term. In MLR, make the same plot against each regressor x_j as well; the plot that shows the pattern tells you which term is missing. 254
- 8.1 The family-wise type I error rate of (8.1) at $\alpha = 0.05$ per test. The intended 5% level is abandoned almost immediately. 269
- 8.2 Between- versus within-treatment variation for the red-team data of Example 8.2. Solid green segments are the treatment means; the dashed line is the grand mean. $SS_{\text{Treatments}}$ measures the green segments' spread around the dashed line; SS_E measures the points' spread around their own green segment. 273
- 8.3 The $F_{2,12}$ density for Example 8.2. The shaded region above $f_{0.05,2,12} = 3.89$ has area 0.05; the observed $F_0 = 8.00$ falls deep inside it. 274
- 8.4 The two yardsticks on a common scale ($f = 15$ error df, $\alpha = 0.05$). They coincide at $a = 2$; for $a \geq 3$ Tukey's grows with the number of treatments, paying for family-wise error control. 282
- 8.5 Underline diagram for the Tukey analysis. Means joined by a common line are not significantly different at family-wise $\alpha = 0.05$. 284
- 8.6 Residuals against fitted values. Panel (a): the red-team residuals from (8.16) show even spread around zero in all three groups — the equal-variance assumption is credible. Panel (b): a funnel pattern from a different experiment — the spread grows with the mean, and a transformation of the response is needed before ANOVA is trusted. 285
- 8.7 Estimated variance components for the warehouse study: 55% of picking-time variance is a between-site effect. 290
- 8.8 The routing data of Example 8.8 by block. All four policies drift upward through the week — the common day effect. The RCBD subtracts that shared drift, so the policies are compared by their vertical separation, not blurred by the slope. 295
- 8.9 Tukey underline diagram for the routing policies. P2–P4 and P4–P1 share lines (not separated); P3 stands alone as the slowest. 296

- 9.1 One-factor-at-a-time versus a factorial design on the same response surface (contours at 70, 80, 90). OFAT explores two perpendicular slices, settles at a response near 67, and never sees the peak. The factorial grid covers the whole region and reveals that the response keeps rising toward the upper right. 313
- 9.2 The three interaction-plot patterns. (a) Parallel lines: the factors act additively; main effects tell the whole story. (b) Lines that diverge without crossing: the direction of the A effect is stable but its size depends on B ; a moderate interaction, common in real data. (c) Crossing lines: which level of A is best reverses with B ; a strong interaction, and main effects alone are misleading. 314
- 9.3 Cell-mean plot for truck type $\times$ terrain. The lines cross between hilly and mountain terrain. 315
- 9.4 Cell means for Example 9.3. The lines are roughly parallel (the rain gap is within noise: $F_0^{AB} = 2.67$, not significant), so main effects summarize the study fairly. 323
- 9.5 Interaction plot for the picking study (Example 9.4). The lines are far from parallel and the layout ranking changes across shifts: layout 1 leads in the evening, layout 2 at night. This is the picture behind $F_0^{AB} = 3.56$. 326
- 9.6 Residual diagnostics for the picking study, following Procedure 9.2. (a) The normal plot is acceptably straight. (b) No funnel against fitted values. (c), (d) Spread is broadly similar across factor levels, with the layout-1 morning cell contributing the largest residuals. 329
- 9.7 The 2^3 design as the corners of a cube: each axis is one factor, and each treatment label marks which factors are high. 334
- 9.8 Normal probability plot of the 15 effect estimates from a single-replicate 2^4 experiment. The ten effects on the dashed line (green circles) behave like noise and are pooled as error; the five that depart from it (A , D , AD , C , AC ; maroon squares) are the active effects. Note that the active set contains interactions, which no OFAT program would have found. 337
- 10.1 The statistical software workflow. Judgment flows in; arithmetic flows back out. 357
- 10.2 The exact two-sided p -value for the yield test: the two shaded tails of the t_9 density beyond ± 2.47 sum to 0.035. The table can only bracket this area between printed columns. 362

- 10.3 An annotated Minitab regression output. Every printed field is an equation from Chapter 6. 366
- 10.4 The four-in-one residual plot for a healthy regression fit: the probability plot is straight (normality), the versus-fits panel is a structureless band (constant variance), the histogram is roughly symmetric, and the versus-order panel shows no drift (independence). 367

List of Tables

- 1.1 Population parameters and the sample statistics that estimate them. Greek letters are unknown targets; the statistics are computed from data. 6
- 1.2 The three basic methods of collecting engineering data. Only the designed experiment supports cause-and-effect conclusions. 9
- 1.3 Miss distances (meters) of a perception module in eight staged pedestrian-crossing scenarios. 12
- 1.4 Frequency distribution of dwell times for 40 import containers. 14
- 1.5 Five-number summaries of time to first booking (minutes) under two on-boarding flows. 16
- 1.6 Key formulas of Chapter 1 at a glance. 21

- 2.1 The parameters studied in this course and their standard point estimators. Each estimator is a statistic, so each has a sampling distribution. 31
- 2.2 Standard normal table excerpt: $\Phi(1.10) = 0.8643$. 40
- 2.3 Standard normal table excerpt: $\Phi(1.50) = 0.9332$. 42
- 2.4 Sampling distributions of the two multi-population estimators. Both are unbiased, and in both the variances of the two samples add. 51
- 2.5 Standard normal table excerpt: $\Phi(1.27) = 0.8980$. 52
- 2.6 Key formulas of Chapter 2 at a glance. 53

- 3.1 Common critical values of the standard normal distribution. Two-sided intervals split α between the tails and use $z_{\alpha/2}$; one-sided bounds put all of α in one tail and use z_{α} . 69
- 3.2 Excerpt of the cumulative z table. Searching the body for 0.9750 gives row 1.9 + column 0.06, so $z_{0.025} = 1.96$. 74

- 3.3 Excerpt of the t table, $t_{\alpha,\nu}$ with $P(T > t_{\alpha,\nu}) = \alpha$. For a 95% two-sided CI with $n = 12$: row $\nu = 11$, column 0.025, giving $t_{0.025,11} = 2.201$. 77
- 3.4 Excerpt of the chi-square table, $\chi^2_{\alpha,\nu}$ with $P(X^2 > \chi^2_{\alpha,\nu}) = \alpha$. A 95% CI for σ^2 with $n = 15$ uses *both* highlighted values in row $\nu = 14$. 80
- 3.5 Two-sided tolerance-interval factors k for a normal population at 95% confidence. The highlighted cell ($n = 25$, coverage 95%) is used in Example 3.6. 84
- 3.6 Interval selection at a glance. The wording of the question picks the row; the available information picks the critical value. 87
- 3.7 Key formulas of Chapter 3 at a glance. All two-sided forms have one-sided versions with z_α or $t_{\alpha,n-1}$. 91
- 4.1 The interval toolkit of Chapter 3 at a glance. Every test in this chapter reuses one of these sampling distributions. 102
- 4.2 The four possible outcomes of a hypothesis test. The two error probabilities α and β are defined in (4.2). 105
- 4.3 Cumulative z-table excerpt: $\Phi(2.50) = 0.9938$, so the upper-tail area beyond 2.50 is 0.0062. 110
- 4.4 Conventional interpretation of p-value magnitudes. 111
- 4.5 Benchmark z values worth memorizing for quick p-value bounds. 112
- 4.6 t -table excerpt. For Example 4.4, $t_0 = 1.40$ falls between the two highlighted entries in the $\nu = 8$ row, so $0.05 < p < 0.10$ one-sided. 116
- 4.7 Decision rules for the chi-square test on a variance (4.7). The distribution is not symmetric, so left and right critical values differ. 117
- 4.8 Chi-square table excerpt. Column headers are *upper-tail* areas: $\chi^2_{0.95,19} = 10.117$ leaves only 5% in the *left* tail. The highlighted pair brackets $\chi^2_0 = 10.64$ in Example 4.5. 118
- 4.9 The four one-sample tests. The sampling distributions are exactly those behind the corresponding confidence intervals in Table 4.1. 123
- 4.10 The three practice problems side by side. The seven-step procedure is identical; only the statistic and its table change. 126
- 4.11 Key formulas of Chapter 4 at a glance. 127
- 5.1 Decision rules for the two-sample z test. The pooled and Welch t tests of the next sections use the same three rows with t critical values in place of z . 142

- 5.2 Excerpt of the t table around $\nu = 20$. The highlighted cell is $t_{0.025,20} = 2.086$, used in Example 5.2; the neighboring columns bound its P -value. 146
- 5.3 Fuel efficiency (km/L) of the same eight vehicles before and after the tune-up, with per-vehicle differences. 149
- 5.4 Independent-samples versus paired designs. If the problem says “same,” “matched,” or “before/after,” or prints the data in corresponding pairs, use the paired t . 151
- 5.5 Excerpt of the F table, upper 0.025 critical values $f_{0.025,\nu_1,\nu_2}$. Green: $f_{0.025,15,20} = 2.57$ for the test in Example 5.6; gold: the swapped-df value $f_{0.025,20,15} = 2.76$ needed by the interval (5.14). 156
- 5.6 Master table for Major Exam 1: every interval and test from Chapters 3–5. One-sample procedures above the rule, two-sample procedures below it. The Welch CI uses the same standard error as its test with $t_{\alpha/2,\nu}$. 159
- 5.7 Key formulas of Chapter 5 at a glance. 166
- 6.1 Regression inference reuses the estimate–standard error– t structure of one-sample inference. Only the formulas for the estimate and its standard error change. 191
- 6.2 Computation table for Example 6.3. Building the five columns and totaling them is Step 2 of Procedure 6.1; every later formula in the analysis reads from the bottom row. 193
- 6.3 t table excerpt: $t_{0.025,3} = 3.182$. 197
- 6.4 The ANOVA table for simple linear regression. The df and SS columns each add up: $1 + (n - 2) = n - 1$ and $SS_R + SS_E = SS_T$ by (6.19). 198
- 6.5 Completed ANOVA table for the ad-spend regression (Example 6.6). 199
- 6.6 F table excerpt ($\alpha = 0.05$): $f_{0.05,1,3} = 10.13$. 199
- 6.7 t table excerpt: $t_{0.025,8} = 2.306$. 202
- 6.8 Confidence interval on the mean response versus prediction interval for a new observation. Exam wording: “mean response” or “average” calls for (6.26); “predict”, “the next”, or “a new” observation calls for (6.27). 204
- 6.9 Residuals and standardized residuals for Example 6.9. 208
- 6.10 Common transformations. After transforming, every formula of this chapter applies to the primed variables unchanged. 209
- 6.11 Key formulas of Chapter 6 at a glance. 213

7.1	Benchmark scores for eight evaluation runs with coded model size and data quality.	232
7.2	The ANOVA table for multiple linear regression. In SLR, $k = 1$ and the residual df reduce to $n - 2$; the structure is otherwise identical.	236
7.4	Excerpt of the F table, $f_{0.05, \nu_1, \nu_2}$. The highlighted cell is $f_{0.05, 2, 5} = 5.79$ used in Example 7.3.	238
7.3	ANOVA table for the evaluation-score model of Example 7.1.	238
7.5	ANOVA table for the delivery-time model, $n = 25$ routes.	240
7.6	Coefficient summary for the delivery-time model. Each t -Stat is Coef divided by SE; each p -value refers to t_{21} .	241
7.7	Excerpt of the t table. The highlighted cell, $t_{0.025, 5} = 2.571$, serves the evaluation model ($n - p = 5$); the rows for 21 and 26 df serve the delivery and churn models.	243
7.8	Coefficient summary for the churn model, $n = 30$ cohorts, 26 residual df.	243
7.9	Choosing the right hypothesis test in multiple regression.	244
7.10	Key formulas of Chapter 7 at a glance.	255
8.1	Dot notation for the single-factor experiment. A dot replaces a subscript that has been summed over; a bar denotes an average.	271
8.2	The one-way ANOVA table. Both the SS column and the df column add down to the Total row — always verify both additions before reading off F_0 .	275
8.3	Red-team defect-discovery data: $a = 3$ strategies, $n = 5$ sessions each, $N = 15$.	276
8.4	Excerpt of the F table, upper 5% points $f_{0.05, \nu_1, \nu_2}$: $f_{0.05, 2, 12} = 3.89$.	277
8.5	Completed ANOVA table for the red-team study. Compare with Figure 8.3.	277
8.6	Onboarding activation-rate data: $a = 3$ flows, $n = 6$ cohorts each, $N = 18$.	279
8.7	ANOVA table for the onboarding experiment. $F_0 = 12.53$ exceeds $f_{0.05, 2, 15} = 3.68$, so the pairwise analysis proceeds.	280
8.8	Excerpt of the t table, upper-tail points $t_{\alpha, \nu}$: $t_{0.025, 15} = 2.131$.	281
8.9	Excerpt of the Studentized range table, upper 5% points $q_{0.05}(a, f)$: $q_{0.05}(3, 15) = 3.67$.	283
8.10	LSD and Tukey HSD applied to the same onboarding data. The borderline A-B gap survives the per-pair yardstick but not the family-wise one.	283

- 8.11 Sample-size search for the guardrail study. Power values are read from the operating characteristic curves for $a - 1 = 3$ numerator degrees of freedom at $\alpha = 0.05$. 287
- 8.12 ANOVA table for the warehouse picking-time study ($a = 4$ randomly sampled warehouses, $n = 5$ orders each). 289
- 8.13 The RCBD ANOVA table. The block row absorbs nuisance variability that would otherwise inflate the error row. 292
- 8.14 Delivery times (minutes per package) for four routing policies blocked by day of week. The last cell is the grand mean $\bar{y}_{..} = 42.0$. 294
- 8.15 RCBD ANOVA table for the routing study. Compare the error mean square with the unblocked value computed in Step 4. 295
- 8.16 RCBD residuals (8.27) for the routing study. Zero margins are automatic; patterns within rows or columns would signal a treatment-by-block interaction. 297
- 8.17 Choosing the right single-factor design. The decision is made when the experiment is planned, before any data exist. 297
- 8.18 Key formulas of Chapter 8 at a glance. 298
-
- 9.1 The ANOVA table for a two-factor factorial with n replicates per cell. All three F ratios share the denominator MS_E . 319
- 9.2 Missed detections per 1,000-frame batch: model version $\times$ scenario type, $n = 2$ batches per cell. Cell totals are $y_{11.} = 18, y_{12.} = 30, y_{13.} = 44, y_{21.} = 16, y_{22.} = 20, y_{23.} = 42$. 321
- 9.3 Excerpt of the $f_{0.05, \nu_1, \nu_2}$ table. The tests in Example 9.3 use the $\nu_2 = 6$ row: 5.99 for one numerator df and 5.14 for two. 322
- 9.4 ANOVA for the perception-model study of Example 9.3. 323
- 9.5 Picking time (seconds per order line), 3×3 factorial with $n = 4$ replicates per cell, with cell, row, and column totals. 324
- 9.6 ANOVA for the picking study. All three effects are significant at $\alpha = 0.05$; the interaction test is read first. 325
- 9.7 Cell means (seconds per order line) for the picking study. 325
- 9.8 Degrees of freedom for a three-factor factorial with n replicates. Each interaction df is the product of the main-effect df of the factors it involves. 330
- 9.9 Treatment-combination notation for the 2^2 design. A lowercase letter appears when that factor is high; (1) means no factor is high. Each label also denotes the total of the n replicates at that combination. 331

9.10	The sign table for the 2^3 design. Main-effect columns code the factor levels; each interaction column is the entrywise product of its parent columns. Reading down the AC column, for example, gives $\text{Contrast}_{AC} = (1) - a + b - ab - c + ac - bc + abc$.	334
9.11	Key formulas of Chapter 9 at a glance.	341
10.1	The statistical software landscape. This course uses Minitab and Python.	355
10.2	The division of labor between the engineer and the software.	356
10.3	Method-to-tool map for the confidence intervals of Chapter 3. All Minitab paths start from Stat $\rightarrow$ Basic Statistics.	359
10.4	Key functions in <code>scipy.stats</code> for the tests of Chapters 4 and 5. (<code>proportions_ztest</code> is in the companion package <code>statsmodels</code> .)	362
10.5	The two-sample menu map. Every path starts from Stat $\rightarrow$ Basic Statistics.	364
10.6	Method-to-tool map for the designed experiments of Chapters 8 and 9.	370
10.7	The complete method-to-tool map for ISE 315. Minitab paths start from the Stat menu; Python functions are in <code>scipy.stats</code> except <code>proportions_ztest</code> , <code>ols</code> , and <code>anova_lm</code> , which are in <code>statsmodels</code> , and the NumPy estimators in the first row.	373
10.8	Software output fields at a glance: what each printed number is, in the notation of this book.	377
A.1	Chapter 1 formulas: descriptive statistics.	442
A.2	Chapter 2 formulas: estimators and sampling distributions.	443
A.3	Chapter 3 formulas: confidence, prediction, and tolerance intervals for a single sample.	445
A.4	Chapter 4 formulas: one-sample hypothesis tests, P-values, and power.	447
A.5	Chapter 5 formulas: two-sample inference.	449
A.6	Chapter 6 formulas: simple linear regression.	451
A.7	Chapter 7 formulas: multiple linear regression. Here C_{jj} is the j th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$.	453
A.8	Chapter 8 formulas: one-way ANOVA, multiple comparisons, and randomized complete block designs.	455
A.9	Chapter 9 formulas: two-factor ANOVA and 2^k factorial designs.	456

- B.1 Cumulative standard normal distribution: $\Phi(z) = P(Z \leq z)$ for $z = -3.49$ to 0.00. Row = sign, integer, and first decimal of z ; column = second decimal. Continues in Table B.2. 459
- B.2 Cumulative standard normal distribution (continued): $\Phi(z) = P(Z \leq z)$ for $z = 0.00$ to 3.49. 460
- B.3 Critical values $t_{\alpha,\nu}$ of the t distribution: $P(T_\nu > t_{\alpha,\nu}) = \alpha$. The last row ($\nu = \infty$) reproduces the standard normal critical values z_α . 461
- B.4 Critical values $\chi_{\alpha,\nu}^2$ of the chi-square distribution: $P(\chi_\nu^2 > \chi_{\alpha,\nu}^2) = \alpha$. Columns 0.995–0.90 give left-tail (small) critical points; columns 0.10–0.005 give right-tail (large) critical points. 462
- B.5 Critical values $f_{0.05,\nu_1,\nu_2}$ of the F distribution: $P(F_{\nu_1,\nu_2} > f_{0.05,\nu_1,\nu_2}) = 0.05$. 463
- B.6 Critical values $f_{0.025,\nu_1,\nu_2}$ of the F distribution: $P(F_{\nu_1,\nu_2} > f_{0.025,\nu_1,\nu_2}) = 0.025$. 464

