

# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Why this textbook? . . . . .	13
1.2	Types of data . . . . .	16
1.3	Data and models . . . . .	16
1.4	Sequence of topics . . . . .	17
1.5	References . . . . .	19
<b>2</b>	<b>Review of Mathematics</b>	<b>21</b>
2.1	Exponents . . . . .	21
2.2	Inequalities . . . . .	23
2.3	Functions . . . . .	24
	2.3.1 Functions in Statistics . . . . .	25
	2.3.2 Plotting functions using SAS - SAS demo . . . . .	26
2.4	Solving linear equations . . . . .	35
2.5	Roots of equations . . . . .	36
2.6	Calculus . . . . .	37
	2.6.1 Derivatives . . . . .	37
	2.6.2 Function plot with derivative - SAS demo . . . . .	39
	2.6.3 Integrals . . . . .	44
2.7	References . . . . .	46
2.8	Problems . . . . .	47
<b>3</b>	<b>Populations and Statistics</b>	<b>49</b>
3.1	Statistical populations . . . . .	49
3.2	Descriptive statistics and frequency . . . . .	50
	3.2.1 Sample mean . . . . .	51
	3.2.2 Median . . . . .	52
	3.2.3 Sample variance . . . . .	53

3.2.4	Standard deviation . . . . .	53
3.2.5	Coefficient of variation . . . . .	53
3.2.6	Range . . . . .	54
3.2.7	Frequency distributions - SAS demo . . . . .	54
3.2.8	Mode . . . . .	62
3.2.9	Skewness . . . . .	62
3.2.10	Kurtosis . . . . .	65
3.2.11	Development time - SAS demo . . . . .	68
3.2.12	Frequency distributions for categorical data - SAS demo	73
3.3	References . . . . .	77
3.4	Problems . . . . .	78
<b>4</b>	<b>Probability Theory</b>	<b>79</b>
4.1	Probability theory . . . . .	79
4.1.1	Events . . . . .	79
4.1.2	Union, intersection, and complement of events . . . . .	80
4.1.3	Probability distributions . . . . .	84
4.1.4	Probability spaces . . . . .	86
4.1.5	Independence of events . . . . .	86
4.1.6	Conditional probability . . . . .	87
4.1.7	A biological probability distribution . . . . .	88
4.1.8	Bayes theorem . . . . .	91
4.2	References . . . . .	95
4.3	Problems . . . . .	96
<b>5</b>	<b>Discrete Random Variables</b>	<b>97</b>
5.1	Binomial distribution . . . . .	98
5.1.1	Binomial distribution - SAS demo . . . . .	100
5.2	Poisson distribution . . . . .	104
5.2.1	Poisson distribution - SAS demo . . . . .	105
5.3	Negative binomial distribution . . . . .	108
5.3.1	Negative binomial distribution - SAS demo . . . . .	108
5.4	Expected values for discrete distributions . . . . .	112
5.4.1	Variance for discrete distributions . . . . .	114
5.5	Discrete random variables and samples . . . . .	116
5.5.1	Parasitic wasps - SAS demo . . . . .	116
5.5.2	Corn borers - SAS demo . . . . .	124
5.6	Classifying spatial or temporal patterns . . . . .	133

5.7	References . . . . .	136
5.8	Problems . . . . .	137
<b>6</b>	<b>Continuous Random Variables</b>	<b>139</b>
6.1	Uniform distribution . . . . .	140
6.1.1	Random sampling coordinates - SAS demo . . . . .	143
6.2	Normal distribution . . . . .	149
6.2.1	Normal distribution - SAS demo . . . . .	151
6.2.2	Sample calculations - standard normal distribution . . . . .	153
6.2.3	Sample calculations - other normal distributions . . . . .	158
6.3	Expected values and variance for continuous distributions . . . . .	161
6.4	Continuous random variables and samples . . . . .	162
6.4.1	Elytra lengths - SAS demo . . . . .	165
6.4.2	Development time - SAS demo . . . . .	170
6.5	References . . . . .	175
6.6	Problems . . . . .	176
<b>7</b>	<b>Expected Value, Variance, and Samples</b>	<b>177</b>
7.1	Expected value and variance . . . . .	177
7.2	Linear functions and sums - expected value and variance . . . . .	179
7.3	Sample mean - expected value and variance . . . . .	180
7.4	Sample variance - expected value . . . . .	181
7.5	Sample calculations and simulation - SAS demo . . . . .	182
7.6	Central limit theorem . . . . .	190
7.6.1	Central limit theorem - SAS demo . . . . .	190
7.7	Applications of the central limit theorem . . . . .	197
7.8	References . . . . .	198
7.9	Problems . . . . .	199
<b>8</b>	<b>Sampling and Estimation</b>	<b>201</b>
8.1	Random samples . . . . .	201
8.2	Parameter estimation . . . . .	202
8.2.1	Maximum likelihood for Poisson data . . . . .	203
8.2.2	Poisson likelihood function - SAS demo . . . . .	205
8.2.3	Maximum likelihood for normal data . . . . .	210
8.2.4	Normal likelihood function - SAS demo . . . . .	211
8.3	Optimality of maximum likelihood estimates . . . . .	216
8.4	References . . . . .	216

8.5	Problems . . . . .	217
<b>9</b>	<b>Confidence Intervals</b>	<b>219</b>
9.1	Preliminaries to confidence intervals . . . . .	219
9.1.1	Parameters and estimates . . . . .	219
9.1.2	Sampling distributions . . . . .	220
9.2	Confidence intervals . . . . .	225
9.2.1	Confidence intervals for $\mu$ when $\sigma^2$ is known . . . . .	226
9.2.2	Confidence intervals for $\mu$ when $\sigma^2$ is estimated . . . . .	228
9.2.3	Confidence intervals for $\sigma^2$ and $\sigma$ . . . . .	230
9.2.4	Confidence intervals - SAS demo . . . . .	232
9.2.5	Confidence interval size . . . . .	233
9.3	References . . . . .	239
9.4	Problems . . . . .	240
<b>10</b>	<b>Hypothesis Testing</b>	<b>241</b>
10.1	The null and alternative hypotheses . . . . .	241
10.2	Test statistics . . . . .	242
10.3	Acceptance and rejection regions – Type I error . . . . .	243
10.3.1	One-sample $Z$ test - sample calculation . . . . .	246
10.4	$P$ values . . . . .	246
10.5	Type II error and power . . . . .	249
10.6	Summary table . . . . .	254
10.7	One-sample $t$ test . . . . .	255
10.7.1	One-sample $t$ test - sample calculation . . . . .	256
10.7.2	Hypothesis testing - SAS demo . . . . .	257
10.7.3	Power analysis for one-sample $t$ tests - SAS demo . . . . .	260
10.8	One-tailed $t$ test . . . . .	264
10.8.1	One-tailed $t$ test - sample calculation . . . . .	266
10.8.2	One-tailed $t$ test - SAS demo . . . . .	266
10.8.3	One-tailed tests - a warning . . . . .	267
10.9	Confidence intervals and tests . . . . .	268
10.10	Likelihood ratio tests . . . . .	269
10.10.1	Example of a likelihood ratio test . . . . .	269
10.11	References . . . . .	273
10.12	Problems . . . . .	274

<b>11 Analysis of Variance (One-Way)</b>	<b>277</b>
11.1 ANOVA models . . . . .	281
11.1.1 Fixed and random effects . . . . .	281
11.1.2 Fixed effects model . . . . .	282
11.1.3 Random effects model . . . . .	285
11.2 Hypothesis testing for ANOVA . . . . .	287
11.2.1 Sums of squares and mean squares . . . . .	287
11.2.2 $F$ statistic and distribution . . . . .	290
11.2.3 ANOVA tables . . . . .	292
11.2.4 One-way ANOVA for Example 1 - SAS demo . . . . .	295
11.2.5 One-way ANOVA for Example 2 - sample calculation . . . . .	302
11.2.6 One-way ANOVA for Example 2 - SAS demo . . . . .	304
11.3 Maximum likelihood estimates . . . . .	311
11.4 $F$ test as a likelihood ratio test . . . . .	313
11.5 One-way ANOVA and two-sample $t$ tests . . . . .	314
11.5.1 Two-sample $t$ test for Example 1 - SAS demo . . . . .	314
11.6 References . . . . .	319
11.7 Problems . . . . .	320
<b>12 Power Analysis for One-Way ANOVA</b>	<b>323</b>
12.1 Power analysis for one-way ANOVA . . . . .	324
12.2 Power analysis - SAS Demo . . . . .	328
12.3 Power analysis continued - SAS demo . . . . .	332
12.4 Power analysis continued - SAS demo . . . . .	335
12.5 References . . . . .	337
12.6 Problems . . . . .	338
<b>13 Multiple Comparisons</b>	<b>339</b>
13.1 Models for multiple comparisons . . . . .	339
13.2 Error rates in multiple comparisons . . . . .	340
13.3 All pairwise comparisons . . . . .	342
13.3.1 Least significant difference . . . . .	343
13.3.2 Least significant difference - SAS demo . . . . .	345
13.3.3 The Tukey procedure . . . . .	353
13.3.4 Tukey procedure - SAS demo . . . . .	354
13.3.5 Multiple range tests - REGW . . . . .	357
13.3.6 REGW procedure - SAS demo . . . . .	359
13.4 Comparisons with a control - Dunnett procedure . . . . .	361

13.4.1	Dunnett's procedure - SAS demo . . . . .	361
13.5	Bonferroni and Sidak corrections . . . . .	363
13.6	Vascular plant cover - SAS demo . . . . .	365
13.7	False discovery rate procedure . . . . .	374
13.7.1	False discovery rate - SAS demo . . . . .	376
13.8	References . . . . .	379
13.9	Problems . . . . .	380
<b>14</b>	<b>Analysis of Variance (Two-Way)</b>	<b>383</b>
14.1	Random assignment of treatments . . . . .	389
14.1.1	Random assignment of treatments - SAS Demo . . . . .	390
14.2	Two-way fixed effects model . . . . .	393
14.2.1	Factor A effect . . . . .	394
14.2.2	Factor B effect . . . . .	394
14.2.3	Factor A and B effect . . . . .	394
14.2.4	Interaction effect . . . . .	394
14.3	Hypothesis testing for two-way ANOVA . . . . .	398
14.3.1	Sum of squares and mean squares . . . . .	398
14.3.2	ANOVA tables and tests . . . . .	402
14.3.3	Two-way ANOVA for Example 1 - SAS demo . . . . .	405
14.3.4	Two-way ANOVA for Example 2 - SAS demo . . . . .	411
14.3.5	Tests for main effects with interaction . . . . .	416
14.4	Unbalanced designs and two-way ANOVA . . . . .	420
14.5	Two-way ANOVA without replication . . . . .	423
14.5.1	Hypothesis testing . . . . .	423
14.5.2	Two-way ANOVA no replication - SAS demo . . . . .	430
14.6	Randomized block designs . . . . .	436
14.6.1	Randomized block models . . . . .	438
14.6.2	Hypothesis testing and variance components . . . . .	438
14.6.3	Randomized block design - SAS demo . . . . .	439
14.6.4	Likelihood ratio test for the block effect . . . . .	447
14.7	References . . . . .	452
14.8	Problems . . . . .	453
<b>15</b>	<b>Assumptions and Transformations</b>	<b>455</b>
15.1	ANOVA assumptions . . . . .	455
15.1.1	Independence of observations . . . . .	455
15.1.2	Homogeneity of variances . . . . .	456

15.1.3	Normality . . . . .	457
15.1.4	Absence of outliers . . . . .	457
15.1.5	Additivity . . . . .	458
15.2	Variance-stabilizing transformations . . . . .	459
15.3	Residual analysis . . . . .	460
15.3.1	Models, estimates, and predictors . . . . .	461
15.3.2	Predicted and residual values . . . . .	461
15.3.3	Evaluating ANOVA assumptions . . . . .	463
15.3.4	Residual analysis and transformations - SAS demo . . . . .	464
15.3.5	$\arcsin(\sqrt{Y})$ transformation - SAS demo . . . . .	470
15.3.6	Transformations when data are limited . . . . .	478
15.4	References . . . . .	479
<b>16</b>	<b>Nonparametric Tests</b>	<b>481</b>
16.1	Wilcoxon two-sample test . . . . .	485
16.1.1	Wilcoxon test for Example 1 - SAS demo . . . . .	487
16.2	Kruskal-Wallis test . . . . .	494
16.2.1	Kruskal-Wallis test for Example 1 - SAS demo . . . . .	495
16.2.2	Kruskal-Wallis test for Example 2 - SAS demo . . . . .	496
16.3	Kolmogorov-Smirnov test . . . . .	499
16.3.1	Kolmogorov-Smirnov test for Example 1 - SAS demo . . . . .	500
16.4	Randomization tests . . . . .	504
16.4.1	Randomization test for Example 3 - SAS demo . . . . .	506
16.5	Limitations of nonparametric tests . . . . .	513
16.6	References . . . . .	515
16.7	Problems . . . . .	516
<b>17</b>	<b>Linear Regression</b>	<b>517</b>
17.1	Linear regression model . . . . .	521
17.2	Linear regression and likelihood . . . . .	521
17.2.1	Sample calculation - $\hat{\beta}$ , $\hat{\alpha}$ , and $F$ test . . . . .	527
17.3	Confidence and prediction intervals . . . . .	530
17.3.1	Sample calculation - confidence and prediction intervals . . . . .	532
17.4	$R^2$ values . . . . .	534
17.5	Linear regression for Example 1 - SAS demo . . . . .	535
17.6	Assumptions and transformations . . . . .	544
17.6.1	Species-area data - SAS demo . . . . .	545
17.6.2	Population growth rates - SAS demo . . . . .	552

17.7	References . . . . .	558
17.8	Problems . . . . .	559
<b>18</b>	<b>Correlation</b>	<b>561</b>
18.1	Correlation model . . . . .	564
18.2	Correlation and maximum likelihood . . . . .	570
18.2.1	Correlation for Example 1 - SAS demo . . . . .	572
18.2.2	Testing $H_0 : \rho = \rho_0$ - SAS demo . . . . .	577
18.2.3	Correlation for <i>I. setosa</i> , all data - SAS demo . . . . .	578
18.3	Correlation assumptions . . . . .	583
18.4	Nonparametric correlation . . . . .	585
18.4.1	Spearman rank correlation for Example 1 - SAS demo . . . . .	587
18.5	References . . . . .	588
18.6	Problems . . . . .	589
<b>19</b>	<b>More Complex ANOVA Designs</b>	<b>591</b>
19.1	Three-way ANOVA . . . . .	591
19.1.1	Three-way fixed effects model . . . . .	594
19.1.2	Three-way ANOVA for Example 1 - SAS demo . . . . .	595
19.1.3	Tests for main effects with interaction . . . . .	604
19.1.4	Other three-way designs . . . . .	608
19.2	One-way nested ANOVA . . . . .	609
19.2.1	Nested ANOVA models . . . . .	611
19.2.2	Nested ANOVA for Example 2 - SAS demo . . . . .	612
19.3	Analysis of covariance . . . . .	620
19.3.1	ANCOVA model . . . . .	622
19.3.2	ANCOVA for Example 3 - SAS demo . . . . .	622
19.4	References . . . . .	629
19.5	Problems . . . . .	630
<b>20</b>	<b>Methods for Categorical Data</b>	<b>633</b>
20.1	Goodness-of-fit tests . . . . .	635
20.1.1	Goodness-of-fit tests for <i>a</i> categories . . . . .	642
20.1.2	Goodness-of-fit tests with estimated parameters . . . . .	647
20.1.3	Corn borers - SAS demo . . . . .	647
20.2	Tests of independence . . . . .	653
20.2.1	Test of independence - sample calculation . . . . .	655
20.2.2	Test of independence - SAS demo . . . . .	656



20.2.3	Test of independence - SAS demo 2 . . . . .	662
20.3	References . . . . .	668
20.4	Problems . . . . .	669
<b>21</b>	<b>Multiple Regression</b>	<b>671</b>
21.1	Multiple regression model . . . . .	676
21.2	Multiple regression in matrix form . . . . .	676
21.3	Multiple regression and likelihood . . . . .	680
21.4	Tests and confidence intervals for $\beta$ . . . . .	687
21.5	Standardized regression coefficients . . . . .	688
21.6	$R^2$ values . . . . .	688
21.7	Multiple regression for Example 1 - SAS demo . . . . .	689
21.8	Visualizing the multiple regression model . . . . .	697
21.9	Collinearity in multiple regression . . . . .	699
21.10	Multiple regression for Example 2 - SAS demo . . . . .	699
21.11	Power analysis for multiple regression . . . . .	707
21.12	Polynomial regression . . . . .	710
21.13	Population growth experiment - SAS demo . . . . .	710
21.14	Model selection using information criteria . . . . .	717
21.15	Model selection for Example 3 - SAS demo . . . . .	723
21.16	References . . . . .	731
21.17	Problems . . . . .	732
<b>22</b>	<b>Data Sets</b>	<b>737</b>
22.1	Elytra Length . . . . .	738
22.2	Development Time . . . . .	742
22.3	Plant Biomass . . . . .	745
22.4	<i>Anagrus</i> fecundity . . . . .	747
22.5	Fitness of <i>T. dubius</i> . . . . .	755
22.6	<i>Iris</i> flower measurements . . . . .	757
22.7	References . . . . .	759
<b>23</b>	<b>Statistical Tables</b>	<b>761</b>
23.1	Table Z: Probabilities for the standard normal distribution. . .	762
23.2	Table T: Quantiles of the $t$ distribution . . . . .	764
23.3	Table C: Quantiles of the $\chi^2$ distribution . . . . .	767
23.4	Table F: Quantiles of the $F$ distribution . . . . .	770

<b>24 Matrix Programs</b>	<b>781</b>
24.1 Matrix calculations . . . . .	782
24.2 Multiple regression in matrix form . . . . .	783
24.3 References . . . . .	785