

Contents

	Preface	xxiii
ONE	Why Do Multivariate Analysis?	1
	OUR FIRST VISIT TO MARS	2
	DUMMIES ON MARS	6
	SUMMARY	7
	PROBLEM	9
TWO	The First Step: Understanding Simple Linear Regression	11
	MORE ON MARS	11
	The Population Parameters	13
	HOW TO ESTIMATE THE LINE OF MEANS FROM A SAMPLE	16
	The Best Straight Line Through the Data	17
	Variability About the Regression Line	22
	Standard Errors of the Regression Coefficients	23
	HOW CONVINCING IS THE TREND?	26
	Testing the Slope of the Regression Line	27
	Comparing Slopes and Intercepts of Two Regression Lines	28
	Testing the Regression as a Whole	29
	Cell Phone Radiation, Reactive Oxygen Species, and DNA Damage in Human Sperm	34

CONFIDENCE INTERVALS FOR REGRESSION	37
Confidence Interval for the Line of Means	38
Confidence Interval for an Observation.	40
CORRELATION AND CORRELATION COEFFICIENTS	41
The Relationship Between Regression and Correlation	43
DOING REGRESSION AND CORRELATION ANALYSIS WITH A COMPUTER.	44
HEAT EXCHANGE IN GRAY SEALS	45
SUMMARY	51
PROBLEMS	51

THREE Regression with Two or More Independent Variables.	57
WHAT WE REALLY DID ON MARS	58
HOW TO FIT THE BEST PLANE THROUGH A SET OF DATA	61
Computing the Regression Coefficients	62
Variability About the Regression Plane	64
Standard Errors of the Regression Coefficients	64
Muddying the Water: Multicollinearity	66
DOES THE REGRESSION EQUATION DESCRIBE THE DATA?	69
Incremental Sums of Squares and the Order of Entry	72
Relationship to <i>t</i> Tests of Individual Regression Coefficients.	74
THE COEFFICIENT OF DETERMINATION AND THE MULTIPLE CORRELATION COEFFICIENT	75
MORE DUMMIES ON MARS	76
MECHANISMS OF TOXIC SHOCK	78
PROTEIN SYNTHESIS IN NEWBORNS AND ADULTS	81
GENERAL MULTIPLE LINEAR REGRESSION	85
Multiple Regression in Matrix Notation	87
DIABETES, CHOLESTEROL, AND THE TREATMENT OF HIGH BLOOD PRESSURE	89
BABY BIRDS BREATHING IN BURROWS	92
POLYNOMIAL (AND SOME OTHER NONLINEAR) REGRESSIONS	95

Heat Exchange in Gray Seals Revisited	96
Other Nonlinear Regressions	99
INTERACTIONS BETWEEN THE INDEPENDENT VARIABLES	100
How Bacteria Adjust to Living in Salty Environments	101
The Response of Smooth Muscle to Stretching	106
SUMMARY	109
PROBLEMS	110
FOUR Do the Data Fit the Assumptions?	117
ANOTHER TRIP TO MARS	119
LOOKING AT THE RESIDUALS	127
A QUANTITATIVE APPROACH TO RESIDUAL ANALYSIS	129
Standardized Residuals	130
Using Residuals to Test for Normality of the Residuals	133
Leverage	138
Studentized Residuals	142
Cook's Distance	144
WHAT DO YOU DO WITH AN INFLUENTIAL OBSERVATION ONCE YOU HAVE FOUND IT?	150
Problems with the Data	151
Problems with the Model	153
Data Transformations	156
WATER MOVEMENT ACROSS THE PLACENTA	160
CHEAPER CHICKEN FEED	165
HOW THE BODY PROTECTS ITSELF FROM EXCESS ZINC AND COPPER	173
Back to Square One	181
CHEAPER CHICKEN FEED REVISITED: BOOTSTRAPPING AND ROBUST STANDARD ERRORS	186
Bootstrap Standard Errors	187
Robust Standard Errors	189
CLUSTERED DATA	193

AIDS ORPHANS IN UGANDA.....	194
SUMMARY	197
PROBLEMS	198

FIVE Multicollinearity and What to Do About It..... 203

WHERE MULTICOLLINEARITY COMES FROM	204
--	-----

BACK TO MARS.....	206
-------------------	-----

DETECTING AND EVALUATING MULTICOLLINEARITY	211
--	-----

Qualitative Suggestions of Harmful Multicollinearity	212
--	-----

Correlations Among the Independent Variables	212
--	-----

The Variance Inflation Factor	213
-------------------------------------	-----

Auxiliary Regressions	215
-----------------------------	-----

The Correlations of the Regression Coefficients	216
---	-----

THE CONSEQUENCES OF HAVING TWO PUMPS IN ONE HEART	216
---	-----

FIXING THE REGRESSION MODEL	221
-----------------------------------	-----

Centering the Independent Variables.....	222
--	-----

Deleting Predictor Variables	229
------------------------------------	-----

More on Two Pumps in One Heart	230
--------------------------------------	-----

FIXING THE DATA	232
-----------------------	-----

Getting More Data on the Heart.....	233
-------------------------------------	-----

USING PRINCIPAL COMPONENTS TO DIAGNOSE AND TREAT MULTICOLLINEARITY.....	238
--	-----

Standardized Variables, Standardized Regression, and the Correlation Matrix.....	239
---	-----

Principal Components of the Correlation Matrix	244
--	-----

Principal Components to Diagnose Multicollinearity on Mars	247
--	-----

Principal Components and the Heart	247
--	-----

Principal Components Regression	249
---------------------------------------	-----

More Principal Components on Mars	252
---	-----

The Catch.....	254
----------------	-----

Recapitulation	256
----------------------	-----

SUMMARY.....	256
--------------	-----

PROBLEMS.....	258
---------------	-----

SIX	Selecting the "Best" Regression Model	261
	SO WHAT DO YOU DO?	262
	WHAT HAPPENS WHEN THE REGRESSION EQUATION CONTAINS THE WRONG VARIABLES?	262
	WHAT DOES "BEST" MEAN?	267
	The Coefficient of Determination R^2	268
	The Adjusted R^2	268
	The Standard Error of the Estimate $s_{y x}$	269
	Independent Validations of the Model with New Data	270
	The Predicted Residual Error Sum of Squares, PRESS	271
	Bias Due to Model Underspecification and C_p	272
	But What Is "Best"?	276
	SELECTING VARIABLES WITH ALL POSSIBLE SUBSETS REGRESSION	277
	What Determines an Athlete's Time in a Triathlon?	278
	SEQUENTIAL VARIABLE SELECTION TECHNIQUES	283
	Forward Selection	284
	Backward Elimination	286
	Stepwise Regression	287
	Interpreting the Results of Sequential Variable Selection	288
	Another Look at the Triathlon	289
	PREDICTIVE OPTIMISM	290
	SUMMARY	295
	PROBLEMS	296
SEVEN	Missing Data	299
	PREVENTION IS KEY	300
	MISSING DATA MECHANISMS	301
	AD HOC MISSING DATA HANDLING METHODS	304
	Listwise Deletion	304
	Single Imputation: Mean and Regression Imputation	305
	MAXIMUM LIKELIHOOD ESTIMATION WITH COMPLETE DATA	306

Using Maximum Likelihood to Estimate Population Mean and Standard Deviation	307
Maximum Likelihood Regression	312
Putting It All Together: Martian Weights Predicted by Heights and Water Consumption via Maximum Likelihood Estimation	312
REGRESSION ANALYSIS VIA MEANS, VARIANCES, AND COVARIANCES	314
The Multivariate Normal Distribution and Covariance	314
Estimating the Regression Coefficients Based on the Mean and Covariance Matrices	318
Back to Mars	319
MAXIMUM LIKELIHOOD REGRESSION ESTIMATION WITH INCOMPLETE DATA	323
Missing Martians	324
Excess Zinc and Copper and Missing Data	329
Missing Data Mechanisms Revisited: Three Mechanisms for Missing Martians	333
NON-NORMAL DATA AND MAXIMUM LIKELIHOOD ESTIMATION	335
Smoking, Social Networks, and Personality	336
The Multivariate Normality Assumption	338
MULTIPLE IMPUTATION	341
Generate the Multiply-Imputed Data Sets	342
Analyzing the Multiply-Imputed Data Sets	347
Combining Results from Multiply-Imputed Data Sets and Quantifying the Uncertainty Due to Multiple Imputation	349
MULTIPLE IMPUTATION EXTENSIONS AND COMPLICATIONS	354
Number of Imputations	354
How Many and Which Independent Variables Should be Included in the Imputation Process?	355
Should a Dependent Variable with Complete Data be Included in Generating Multiple Imputations?	359
Small Samples	360
Non-Normal Data	360
Clustered Data	361

Data Not Missing at Random	362
Excess Zinc and Copper and Missing Data Revisited.	363
SUMMARY	367
PROBLEMS	369

EIGHT One-Way Analysis of Variance	371
USING A t TEST TO COMPARE TWO GROUPS	372
Does Secondhand Tobacco Smoke Nauseate Martians?	373
Using Linear Regression to Compare Two Groups	376
THE BASICS OF ONE-WAY ANALYSIS OF VARIANCE	378
Traditional Analysis-of-Variance Notation	380
Accounting for All the Variability in the Observations.	384
Expected Mean Squares	387
Using Linear Regression to Do Analysis of Variance with Two Groups	388
USING LINEAR REGRESSION TO DO ONE-WAY ANALYSIS OF VARIANCE WITH ANY NUMBER OF TREATMENTS.	391
Hormones and Depression	393
MULTIPLE COMPARISON TESTING.	399
The Bonferroni t Test	401
More on Hormones and Depression	402
Holm t Test	404
Holm-Sidak t Test	406
What Is a Family?	408
Diet, Drugs, and Atherosclerosis	409
TESTING THE ASSUMPTIONS IN ANALYSIS OF VARIANCE.	412
Formal Tests of Homogeneity of Variance	413
More on Diet, Drugs, and Atherosclerosis.	415
Alternatives to the Usual F Statistic When Variances Are Unequal.	417
Alternatives to the t Test Statistic When Variances Are Unequal.	421
Maturing Rat Lungs	422
SUMMARY	425
PROBLEMS	425

NINE Two-Way Analysis of Variance	429
TRADITIONAL TWO-WAY ANALYSIS OF VARIANCE	432
PERSONALITY ASSESSMENT AND FAKING HIGH GENDER IDENTIFICATION	435
Traditional Analysis of Variance	436
Using Regression to Perform Two-Way Analysis of Variance	439
AN ALTERNATIVE APPROACH FOR CODING DUMMY VARIABLES	447
An Alternative Approach to Personality	448
Why Does It Matter How We Code the Dummy Variables?	451
The Kidney, Sodium, and High Blood Pressure	454
What Do Interactions Tell Us?	459
Multiple Comparisons in Two-Way Analysis of Variance	463
More on the Kidney, Sodium, and High Blood Pressure	464
UNBALANCED DATA	469
ALL CELLS FILLED, BUT SOME CELLS HAVE MISSING OBSERVATIONS	471
The Case of the Missing Kidneys	472
Summary of the Procedure	477
What If You Use the Wrong Sum of Squares?	479
Multiple Comparisons with Missing Data	481
ONE OR MORE CELLS EMPTY	484
Multiple Comparisons with Empty Cells	487
More on the Missing Kidney	490
Multiple Comparisons for the Missing Kidney	494
Recapitulation	496
RANDOMIZED BLOCK DESIGNS	496
A More Thorough Study of Martian Nausea	497
What Do You Gain by Blocking?	500

Regression Implementation of Randomized Blocks	502
Recapitulation	504
SUMMARY	504
PROBLEMS	506
TEN Repeated Measures	511
ONE-WAY REPEATED-MEASURES ANALYSIS OF VARIANCE	513
Hormones and Food	514
Comparison with Simple Analysis of Variance.	520
Multiple Comparisons in Repeated-Measures Analysis of Variance.	521
Recapitulation	522
TWO-FACTOR ANALYSIS OF VARIANCE WITH REPEATED MEASURES ON ONE FACTOR.	522
Partitioning the Variability	524
Testing the Non-Repeated-Measures Factor	525
Testing the Repeated-Measures Factor.	526
Is Alcoholism Associated with a History of Childhood Aggression?	528
THE GENERAL LINEAR MODEL	538
Traditional Regression Model Structure for Mixed Models Estimated by Maximum Likelihood.	547
Maximum Likelihood Estimation for Linear Mixed Models	553
HYPOTHESIS TESTING IN MAXIMUM LIKELIHOOD ESTIMATION	555
Testing the ANOVA Effects Using the Regression Equation	555
Testing the Individual Coefficients.	559
The Wald χ^2 and F Tests	562
Maximum Likelihood Mixed Models Analysis of Drinking and Antisocial Personality Using an Overspecified Model	564
Maximum Likelihood Mixed Models Analyses of the Study of Drinking and Antisocial Personality: Including Subjects with Missing Data on the Dependent Variable	564
BETTER ESTIMATES OF THE COVARIANCE STRUCTURE FOR REPEATED MEASURES	569
Treating Between-Subjects Effects as Random Effects	569
Estimating the Covariance in the Residuals	570
This Is Your Rat's Brain on Drugs	573