

Lecture notes for
Advanced Data Analysis 1 (ADA1)
Stat 427/527
University of New Mexico

Edward J. Bedrick

Ronald M. Schrader

Erik B. Erhardt

Fall 2011

Contents

Course details	vii
Introduction to Statistical Packages	ix
1 Summarizing and Displaying Data	1
1.1 Numerical Summaries	1
1.2 Graphical Summaries	4
1.2.1 Dotplots	4
1.2.2 Histogram and Stem-and-Leaf	5
1.2.3 Boxplots	8
1.3 Interpretation of Graphical Displays for Numerical Data	10
1.4 Interpretations for examples	16
1.5 Minitab Discussion	17
2 Estimation in One-Sample Problems	19
2.1 Inference for a population mean	19
2.2 CI for μ	20
2.2.1 Assumptions for procedures	20
2.2.2 The effect of α on a two-sided CI	21
2.3 Hypothesis Testing for μ	22
2.3.1 Assumptions for procedures	23
2.3.2 P-values	24
2.3.3 The mechanics of setting up hypothesis tests	29
2.3.4 The effect of α on the rejection region of a two-sided test	30
2.4 Two-sided tests, CI and p-values	32
2.5 Statistical versus practical significance	32
2.6 Design issues and power	33
2.7 One-sided tests on μ	33
3 Two-Sample Inferences	39
3.1 Comparing Two Sets of Measurements	39
3.1.1 Salient Features to Notice	43
3.2 Two-Sample Methods: Paired Versus Independent Samples	44
3.3 Two Independent Samples: CI and Test Using Pooled Variance	44
3.4 Satterthwaite's Method	45
3.4.1 MINITAB Implementation	46

3.5	One-Sided Tests	49
3.6	Paired Analysis	50
3.7	Minitab Analysis	50
3.8	Should You Compare Means?	54
4	Checking Assumptions	57
4.1	Testing Normality	57
4.1.1	Formal Tests of Normality	63
4.2	Testing Equal Population Variances	65
5	One-Way Analysis of Variance	67
5.1	Multiple Comparison Methods: Fisher's Method	72
5.1.1	Be Careful with Interpreting Groups in Multiple Comparisons!	74
5.2	FSD Multiple Comparisons in Minitab	74
5.2.1	Discussion of the FSD Method	75
5.3	Bonferroni Comparisons	76
5.4	Further Discussion of Multiple Comparisons	78
5.5	Checking Assumptions in ANOVA Problems	79
5.6	Example from the Child Health and Development Study (CHDS)	81
6	Nonparametric Methods	85
6.1	The Sign Test and CI for a Population Median	85
6.2	Wilcoxon Signed-Rank Procedures	88
6.2.1	Nonparametric Analyses of Paired Data	90
6.2.2	Comments on One-Sample Nonparametric Methods	91
6.3	(Wilcoxon-)Mann-Whitney Two-Sample Procedure	93
6.4	Alternative Analyses for ANOVA and Planned Comparisons	97
6.4.1	Kruskal-Wallis ANOVA	97
6.4.2	Transforming Data	98
6.4.3	Planned Comparisons	104
6.4.4	A Final ANOVA Comment	106
7	Categorical Data Analysis	107
7.1	Single Proportion Problems	107
7.1.1	A CI for p	107
7.1.2	Hypothesis Tests on Proportions	109
7.1.3	The p-value for a two-sided test	110
7.1.4	Appropriateness of Test	111
7.1.5	Minitab Implementation	111
7.1.6	One-Sided Tests and One-Sided Confidence Bounds	112
7.1.7	Small Sample Procedures	113
7.2	Analyzing Raw Data	115
7.3	Goodness-of-Fit Tests (Multinomial)	117
7.3.1	Adequacy of the Goodness-of-Fit Test	119
7.3.2	Minitab Implementation	119
7.3.3	Multiple Comparisons in a Goodness-of-Fit Problem	120

7.4	Comparing Two Proportions: Independent Samples	121
7.4.1	Large Sample CI and Tests for $p_1 - p_2$	122
7.4.2	Minitab Implementation	124
7.5	Effect Measures in Two-by-Two Tables	127
7.6	Analysis of Paired Samples: Dependent Proportions	128
7.7	Testing for Homogeneity of Proportions	130
7.7.1	Minitab Implementation	132
7.8	Testing for Homogeneity in Cross-Sectional and Stratified Studies	134
7.8.1	Testing for Independence in a Two-Way Contingency Table	135
7.8.2	Further Analyses in Two-Way Tables	136
7.9	Adequacy of the Chi-Square Approximation	138
8	Correlation and Regression	141
8.1	Testing that $\rho = 0$	142
8.2	The Spearman Correlation Coefficient	143
8.3	Simple Linear Regression	145
8.3.1	Linear Equation	145
8.3.2	Least Squares	145
8.3.3	Minitab Implementation	147
8.4	ANOVA Table for Regression	147
8.4.1	Brief Discussion of Minitab Output for Blood Loss Problem	150
8.5	The regression model	150
8.5.1	Back to the Data	151
8.6	CI and tests for β_1	151
8.6.1	Testing $\beta_1 = 0$	152
8.7	A CI for the population regression line	152
8.7.1	CI for predictions	153
8.7.2	A further look at the blood loss data (Minitab Output)	154
8.8	Checking the regression model	154
8.8.1	Checking normality	158
8.8.2	Checking independence	158
8.8.3	Outliers	158
8.8.4	Influential observations	159
9	Model Checking and Regression Diagnostics	161
9.1	Residual Analysis	162
9.2	Checking Normality	163
9.3	Checking Independence	164
9.4	Outliers	164
9.5	Influential Observations	164
9.6	Diagnostic Measures available in Minitab	166
9.7	Minitab Regression Analysis	169
9.8	Residual and Diagnostic Analysis of the Blood Loss Data	170
9.9	Gesell data	174
9.10	Weighted Least Squares	178

A	Statistical Tables	181
A.1	Normal (z) distribution	182
A.2	t distribution	183
A.3	χ^2 distribution	184