Statistics

Xi Tan (tan19@purdue.edu)

November 16, 2017

Contents

Preface

This booklet is divided into 7 Chapters. The first chapter introduces the definitions of basic concepts, such as event, sample space, and probability space. Followed in the next chapter, we will discuss the relationship between two or more events when they interplay with each other. The third chapter formally brings in random variables and vectors, as a basis to develop their quantitative measure and characteristic functions later in chapter four. Chapter five includes some well-known limit theorems, which is useful for asymptotic analysis. The last two chapters will discuss several selected topics in probability theory, and provide a summary of common distributions.

Six types of statistical analysis:

1. Descriptive
2. Exploratory
3. Inferential (parameters)
4. Predictive (what will happen)
5. Causal (why it happens)
6. Mechanistic (how to deal with it)

A correlation matrix is defined as:

$$\text{Corr}(\mathbf{X}) = (\text{diag}(\mathbf{\Sigma}))^{-\frac{1}{2}}\mathbf{\Sigma}(\text{diag}(\mathbf{\Sigma}))^{-\frac{1}{2}} \tag{1}$$

Both covariance matrices and correlation matrices are symmetric semipositive definite matrices.

Main References:

1. Extending the Linear Model with R, by Julian J. Faraway
2. Categorical Data Analysis 3rd Edition, by Alan Agresti
3. Generalized, Linear, and Mixed Models, by Charles E. McCulloch, Shayle R. Searle, John M. Neuhaus
4. An Introduction to Generalized Linear Models, by Annette J. Dobson, Adrian Barnett
5. Generalized Linear Models, by P. McCullagh, John A. Nelder

Part I

Statistical Models

Chapter 1

Collecting Data: Experiments and Surveys

1.1 Design of Experiments

1.2 Statistical Survey

1.3 Opinion Poll

1.4 Sampling

Sampling Distribution

Sampling: Stratified Sampling, Quota Sampling

Biased Sample: Spectrum Bias, Survivorship Bias

Chapter 2

Describing Data

Chapter 3

Filtering Data

3.1 Recursive Bayesian Estimation

Kalman Filter

Particle Filter

3.2 Moving Average

Chapter 4

Linear Regression Models

4.1 Introduction

Generalized linear models include as special cases, linear regression and analysis-of-variance models, logit and probit models for quantal responses, log linear models and multinomial response models for counts and some commonly used models for survival data.

The second-order properties of the parameter estimates are insensitive to the assumed distributional form: the second-order properties depend mainly on the assumed variance-to-mean relationship and on uncorrelatedness or independence.

Data types:

4.2 Simple Linear Regression

Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{4.1}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Estimated Regression Function

$$b_1 = \rho_{XY} \cdot \frac{s_Y}{s_X} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{i=1}^n \left[\frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] Y_i \tag{4.2}$$

$$b_0 = \bar{Y} - b_1 \bar{X} \tag{4.3}$$

$$\hat{\sigma}^2 = \frac{MSE}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} \tag{4.4}$$

Notice, $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$, and $b_1 = \rho \cdot \frac{s_Y}{s_X}$, where ρ is the correlation between X and Y and s_Y, s_X are standard error of Y and X, respectively.

The slope of the fitted line is equal to the correlation between y and x corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass $(\bar{x}, \bar{y})$ of the data points.

Another way of writing the estimated regression function is

$$\hat{Y}_i = \bar{Y} + b_1 (X_i - \bar{X}) \tag{4.5}$$

Notice, $\bar{Y}$ and b_1 are uncorrelated (check it using the fact that $b_1 = \sum_{i=1}^n k_i Y_i$).

Inference About b_1 and b_0

Since $SSE/\sigma^2 \sim \chi^2_{n-2}$, and $\frac{s^2\{b_1\}}{\sigma^2\{b_1\}} \sim \frac{\chi^2_{n-2}}{n-2}$

$$\frac{b_1-\beta_1}{s\{b_1\}} = \frac{b_1-\beta_1}{\sigma\{b_1\}} \Big/ \frac{s\{b_1\}}{\sigma\{b_1\}} \sim \frac{z}{\sqrt{\frac{\chi^2_{n-2}}{n-2}}} = t_{n-2} \tag{4.6}$$

so the confidence interval for b_1, with confidence level α is

$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\} \tag{4.7}$$

or

$$b_1 \mp t(\alpha/2; n-2)s\{b_1\} \tag{4.8}$$

Similarly, the confidence interval for b_0, with confidence level α is

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\} \tag{4.9}$$

or

$$b_0 \mp t(\alpha/2; n-2)s\{b_0\} \tag{4.10}$$

The power of testing $\beta_1 = \beta^{H_0}$ is $Power = P\{|t^*| > t(1-\alpha/2; n-2)|\delta\}$, where $\delta = \frac{|\beta_1-\beta^{H_0}|}{\sigma\{b_1\}}$. Similar for β_0.

Properties of k_i

$$k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i-\bar{X})^2} \tag{4.11}$$

$$\sum_{i=1}^n k_i = 0 \tag{4.12}$$

$$\sum_{i=1}^n k_i X_i = 1 \tag{4.13}$$

$$\sum_{i=1}^n k_i^2 = \frac{1}{\sum_{i=1}^n (X_i-\bar{X})^2} \tag{4.14}$$

The second and third identities hold as a requirement for the unbiasness, since

$$E(b_1) = E\left(\sum k_i Y_i\right) = E\left(\sum k_i(\beta_0+\beta_1 X_i)\right) = E\left(k_i \sum \beta_0 + \beta_1 \sum k_i X_i\right) = \beta_1$$

requires $\sum k_i = 0$ and $\sum X_i k_i = 1$. The fourth identity ensures the attainment of the minimum variance.

Properties of e_i

$$e_i = Y_i - \hat{Y}_i \tag{4.15}$$

$$\sum e_i = 0 \tag{4.16}$$

$$\sum X_i e_i = 0 \tag{4.17}$$

$$\sum \hat{Y}_i e_i = 0 \tag{4.18}$$

Properties of b_1 and b_0

$$b_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum(X_i - \bar{X})^2}\right) \tag{4.19}$$

$$b_0 \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum(X_i - \bar{X})^2}\right) \tag{4.20}$$

where σ^2 can be estimated by the MSE, i.e., $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$

Now, since

$$b_1 = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sum_{i=1}^{n}(X_i - \bar{X})Y_i \tag{4.21}$$

$$= \sum_{i=1}^{n} k_i Y_i \tag{4.22}$$

where $k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$, we have

$$\sum k_i = 0 \tag{4.23}$$

$$\sum X_i k_i = 1 \tag{4.24}$$

$$\sum k_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{4.25}$$

The first two identity hold as a requirement for the unbiasness, since

$$E(b_1) = E\left(\sum k_i Y_i\right) = E\left(\sum k_i(\beta_0 + \beta_1 X_i)\right) = E\left(\beta_0 \sum k_i \beta_0 + \beta_1 \sum k_i X_i\right) = \beta_1$$

requires $\sum k_i = 0$ and $\sum X_i k_i = 1$. The third identity ensures the attainment of the minimum variance.

	Estimate	Expectation	Variance
Y_i	$\hat{Y}_i$	$\beta_0 + \beta_1 X_i$	σ^2
b_1	$\frac{\sum(X_i-\bar{X})Y_i}{\sum(X_i-\bar{X})^2}$	β_1	$\sigma^2 \cdot \frac{1}{\sum(X_i-\bar{X})^2}$
b_0	$\bar{Y} - b_1\bar{X}$	β_0	$\sigma^2 \cdot \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right]$
$\hat{Y}_h$	$\bar{Y} + b_1(X_h - \bar{X})$	$\beta_0 + \beta_1 X_h$	$\sigma^2 \cdot \left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]$
$\hat{Y}_{h(new)}$	$\bar{Y} + b_1(X_h - \bar{X})$	$\beta_0 + \beta_1 X_h$	$\sigma^2 \cdot \left[1 + \frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]$
$\hat{Y}_{h(new_m)}$	$\bar{Y} + b_1(X_h - \bar{X})$	$\beta_0 + \beta_1 X_h$	$\sigma^2 \cdot \left[\frac{1}{m} + \frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]$
e_i	$Y_i - \hat{Y}_i$	0	$1 - h_{ii}$

Table 4.1: Simple Linear Regression

In particular, when $X_h = 0$ we obtain the formulas for b_0, and when $X_h - \bar{X} = 1$ we obtain the formulas for b_1.

ANOVA of Simple Linear Regression Model

$$SSTO = SSR + SSE \tag{4.26}$$

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \sum_{i=1}^{n}(\hat{Y}_i - \bar{y}) + \sum_{i=1}^{n}(\bar{y} - \hat{Y}_i) \tag{4.27}$$

SSR can also be computed as $SSR = b_1^2 \sum_{i=1}^{n}(X_i - \bar{X})$, so given the same "distribution" of X, the steeper the slope of the regression line, the higher the SSR, and hence the better fit of the model.

To test $H_0 : \beta_1 = 0$, we use $F = \frac{SSR}{SSE}$. There is equivalence between an F test and a t test: $[t(1-\alpha/2, n-2)]^2 = F(1-\alpha, n-2)$.

Chapter 5

Generalized Linear Regression Models

5.1 Survival Analysis

$$S(t) = \exp\left[-\int_0^t \lambda(u)du\right] \tag{5.1}$$

$$L(\lambda) = \prod_{i=1}^{n}[\lambda(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i} \tag{5.2}$$

where $S(t)$ is the survival function, and $\lambda(t)$ is the hazard function.

	Estimate	Standard Error	NOTE
S	$\hat{S(t)} = \prod \frac{n_j - d_j}{n_j}$	$\hat{S(t)}\sqrt{\sum \frac{d_i}{n_j(n_j-d_j)}}$	
Λ	$-\log \hat{S(t)}$	$\sqrt{\sum \frac{d_i}{n_j(n_j-d_j)}}$	
λ	$\frac{\sum \delta_i}{\sum(X_i - V_i)}$	$\frac{\hat{\lambda}}{\sqrt{\sum \delta_i}}$	

Table 5.1: Survival Analysis

Chapter 6

Analysis of Variance (ANOVA)

Chapter 7

Multivariate Analysis

7.1 Principal Component Analysis (PCA)

ALgebraically, principal components are particular linear combinations of the p random variables $X_1, \cdots, X_p$. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, \cdots, X_p$ as the coordinate axes.

7.2 Factor Analysis

7.3 Cluster Analysis

7.4 Discriminant Analysis

7.5 Correspondence Analysis

7.6 Canonical Correlation Analysis (CCA)

7.7 Multidimensional Scaling (MDS)

Chapter 8

Modeling Sample Data

8.1 Density Estimation

Kernel Density Estimation

Multivariate Kernel Density Estimation

8.2 Time Series

8.3 Robust Statistics

Chapter 9

Modeling Population Data: Statistical Inference

9.1 Bayesian Inference

Bayes' theorem, Bayes Estimator, Prior Distribution, Posterior Distribution, Conjugate Prior, and All That

9.2 Frequentist Inference

Statistical Hypothesis Testing: Null, Alternative, P-value, Significance level, power, likelihood-ratio test, goodness-of-fit, confidence interval, M-estimator, Trimmed Estimator

9.3 Non-parametric Statistics

Nonparametric Regression, Kernel Methods

Chapter 10

Making Decisions: Decision Theory

10.1 Optimal Decision, Type I and Type II errors

10.2 Correlation and Causation

When a statistical test shows a correlation between A and B, there are usually five possibilities:

1. A causes B.
2. B causes A.
3. A and B both partly cause each other.
4. A and B are both caused by a third factor, C.
5. The observed correlation was due purely to chance.

Chapter 11

Theory of Linear Models

11.1 Linear Models, Estimable Functions, Least Squares Estimates=LSE, Normal Equations, Projections, Gauss Markov theorem, BLUE

11.2 Multivariate Normal Distribution and Distribution of Linear and Quadratic Forms

11.3 Properties of LSE and Generalized LSE

11.4 General Linear Hypothesis=GLH, Testing of GLH

11.5 Orthogonalization of Design Matrix and Canonical Reduction of GLH; Adding Variables To The Model

11.6 Correlation, Multiple Correlation and Partial Correlation

11.7 Confidence Regions and Prediction Regions

11.8 Simultaneous Confidence Sets, Bonferroni, Scheffe Projection Method, Tukey Studentized Range

11.9 Introduction to Design of Experiments, ANOVA and ANOCOVA, Factorial and Block Designs, Random, Fixed and Mixed Models, Components of Variance

11.10 Hierarchial Bayes Analysis of Variance; (Schervish Ch. 8, 8.1,8.2) Partial Exchangeability and Hierarchical, Models, Examples and Representations, Normal One Way ANOVA and Two Way Mixed Model ANOVA

Part II

Mathematical Statistics

Chapter 12

Introduction

12.1 Degrees of Freedom

12.2 sufficient, complete, and etc.

12.3 Likelihood Function

12.4 Exponential Family

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \tag{12.1}$$

$$E(y) = b'(\theta) \tag{12.2}$$

$$Var(y) = b''(\theta)a(\phi) \tag{12.3}$$

12.5 Cramer-Rao Theorem

Chapter 13

Data, Models, Statistics, Parameters

13.1 Distributions of Functions of a Random Variable

Theorem 13.1.1 From Casella & Berger theorem 2.1.5) Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous and that $g^{-1}(y)$ has a continuous derivative. Then the pdf of Y is given by

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| \tag{13.1}$$

Chapter 14

Decision Theory (Bayes and Minimax Criteria, Risk Functions, Estimation and Testing in Terms of the Decision Theoretic Framework

Chapter 15

Bayesian Models, Conjugate (and Other) Prior Distributions

Chapter 16

Prediction (Optimal MSPE and Optimal Linear MSPE)

Chapter 17

Sufficiency (Factorization theorem)

Chapter 18

Natural Sufficient Statistics

Chapter 19

Minimal Sufficiency

Chapter 20

Estimation (Least Squares, MLE, Frequency Plug-in, Method of Moments, Combinations of These)

Chapter 21

Exponential Families & Properties, Canonical Exponential Families (& Fisher Information)

Chapter 22

Information Inequality, Fisher Information, UMVU Estimates, Cramer-Rao Lower Bound

Chapter 23

Neyman-Pearson Testing Theory (Form, MP Test, UMP Test, MLR Family, Likelihood Ratio Tests)

Chapter 24

Asymptotic Approximation / Large Sample Theory (Consistency, Delta Method, Asymptotic Normality of MLE, Slutsky's theorem, Efficiency, Pearson's Chi-Square)

Chapter 25

Selected Topics

25.1 M-estimator

25.2 Sweep Operator

25.3 Information Geometry

25.4 Boostrap

Part III

Appendix