SPRINGER BRIEFS IN STATISTICS

George A. F. Seber

Statistical Models for Proportions and Probabilities

Springer

SpringerBriefs in Statistics

For further volumes:
http://www.springer.com/series/8921

George A. F. Seber

Statistical Models for Proportions and Probabilities

Springer

George A. F. Seber
Department of Statistics
The University of Auckland
Auckland
New Zealand

ISSN 2191-544X ISSN 2191-5458 (electronic)
ISBN 978-3-642-39040-1 ISBN 978-3-642-39041-8 (eBook)
DOI 10.1007/978-3-642-39041-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013942014

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Most elementary statistics books discuss inference for proportions and probabilities, and the primary readership for this monograph is the student of statistics, either at an advanced undergraduate or graduate level. As some of the recommended so-called "large-sample" rules in textbooks have been found to be inappropriate, this monograph endeavors to provide more up-to-date information on these topics. I have also included a number of related topics not generally found in textbooks. The emphasis is on model building and the estimation of parameters from the models.

It is assumed that the reader has a background in statistical theory and inference and is familiar with standard univariate and multivariate distributions, including conditional distributions. This monograph may also be helpful for the statistics practitioner who is involved with statistical consulting in this area, particularly with regard to inference for one and two proportions or probabilities.

Chapter 1 looks at the difference between a proportion and probability. It focuses on a proportion leading to the Hypergeometric model and its Binomial approximation, along with inference for the proportion. Inverse sampling is also considered. Chapter 2 focuses on estimating a probability and considers the Binomial distribution in detail as well as inverse sampling. Exact and approximate inferences for a probability are considered. In Chap. 3, the main focus is on comparing two proportions or two probabilities and related quantities such as the relative risk and the odds ratio from the same or different populations using the Multi-hypergeometric or Multinomial distributions. Simultaneous confidence intervals for several parameters are also considered. The Multinomial distribution is the basis for a number of hypothesis and goodness of fit tests, and these are discussed in Chap. 4 with particular attention given to 2×2 tables and matched data. In Chap. 5, we look briefly at two logarithmic models for discrete data, namely the log linear and the logistic models.

I would like to thank two reviewers for their very helpful comments on a previous draft.

Auckland, New Zealand, June 2012 George A. F. Seber

Contents

Chapter 1
Single Proportion

Abstract This chapter focusses on the problem of estimating a population proportion using random sampling with or without replacement, or inverse sampling. Exact and approximate confidence intervals are discussed using the Hypergeometric distribution. Applications to capture-recapture models are given.

Keywords Proportion · Hypergeometric distribution · Simple random sample · Sampling fraction · Negative-Hypergeometric distribution · Binomial distribution · Confidence intervals for a proportion · Single capture-recapture model

1.1 Distribution Theory

Since this monograph is about modelling proportions and probabilities, I want to begin by comparing the two concepts, proportion and probability, as these two ideas are sometimes confused. If we have a population of N people and M are male, then the proportion of males in the population is $p = M/N$. Suppose we now carry out a random experiment and choose a person at random from the population. What we mean by this is that we choose a person in such a way that every person is equally likely to be chosen. If the population is small we could achieve this by putting the names of everyone in a container, shuffling the names by rotating the container, and drawing one name out. This kind of manual process is used in lottery games. For example in New Zealand we have Lotto in which 40 numbered balls are tossed around in a container until one eventually drops out.

For a large population of people we could number everyone and then choose a number at random using a computer. In this case we can obtain the probability of getting a male using the law of probability relating to equally likely events, namely if we have N equally likely outcomes of an experiment (the so-called sample space) and M of these have a given characteristic, then the probability of choosing a member with the given characteristic is simply M/N or p again. This means that for the

G. A. F. Seber, *Statistical Models for Proportions and Probabilities*, SpringerBriefs in Statistics, DOI: 10.1007/978-3-642-39041-8_1,