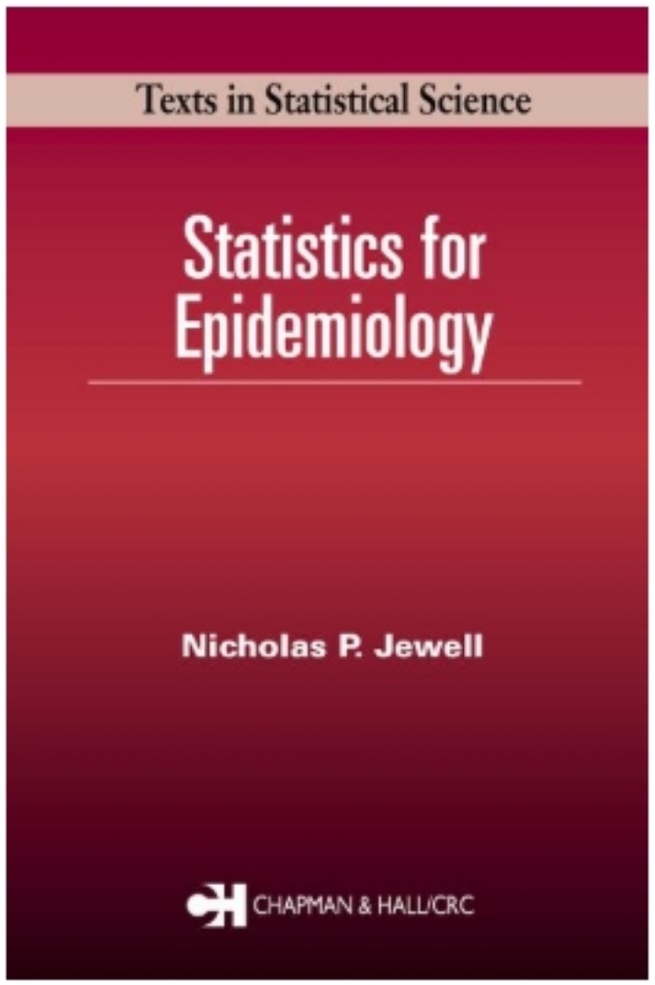

Statistics for Epidemiology

1st Edition

Nicholas P. Jewell

Chapman and Hall/CRC
Published August 26, 2003
Textbook - 350 Pages - 47 B/W Illustrations
ISBN 9781584884330 - CAT# C4339
Series: Chapman & Hall/CRC Texts in Statistical Science

Summary

Statistical ideas have been integral to the development of epidemiology and continue to provide the tools needed to interpret epidemiological studies. Although epidemiologists do not need a highly mathematical background in statistical theory to conduct and interpret such studies, they do need more than an encyclopedia of "recipes."

Statistics for Epidemiology achieves just the right balance between the two approaches, building an intuitive understanding of the methods most important to

Texts in Statistical Science

Statistics for Epidemiology

Nicholas P. Jewell

CHAPMAN & HALL/CRC

Statistics for Epidemiology

Nicholas P. Jewell

CHAPMAN & HALL/CRC
A CRC Press Company
Boca Raton London NewYork Washington, D.C.

This edition published in the Taylor & Francis e-Library, 2009.

To purchase your own copy of this or any of
Taylor & Francis or Routledge's collection of thousands of eBooks
please go to www.eBookstore.tandf.co.uk.

Datasets and solutions to exercises can be downloaded at http://www.crcpress.com/e_products/downloads/.

Send correspondence to Nicholas P.Jewell, Division of Biostatistics, School of Public Health, 140 Warren Hall #7360, University of California, Berkeley, CA 94720, USA. Phone: 510-642-4627, Fax: 510-643-5163, e-mail: jewell@stat.berkeley.edu

Library of Congress Cataloging-in-Publication Data

Statistics for epidemiology/by Nicholas P.Jewell.
p.cm.—(Texts in statistical science series; 58)
Includes bibliographical references and index.
ISBN 1-58488-433-9 (alk. paper)
1. Epidemiology—Statistical methods. I. Jewell, Nicholas P, 1952– II. Texts in statistical science.
RA652.2.M3S745 2003
614.4′072′7–dc21 2003051458

ISBN 0-203-49686-8 Master e-book ISBN

ISBN 0-203-59461-4 (Adobe ebook Reader Format)
International Standard Book Number 1-58488-433-9
Library of Congress Card Number 2003051458

1. Describe a scenario where a χ^2 test for trend will have less power to detect an association between beef consumption and the disease than the overall χ^2 test of association.

2. Suppose you calculate both χ^2_{trend} and $\chi^2_{overall}$. The goodness of fit statistic $\chi^2_{gof} = \chi^2_{overall} - \chi^2_{trend}$ is calculated to be 13.8. How many degrees of freedom should be used for the reference χ^2 distribution? What is your interpretation of the observed χ^2_{gof}?

Question 11.2

The data on biomass cooking fuel exposure and tuberculosis (Perez-Padilla et al., 2001) (see Question 6.4) included additional information on age of the respondent. Table 11.4 provides the data on exposure and tuberculosis across various age subgroups. For example, among respondents less than 25 years old, there were 59 tuberculosis cases, of whom 8 were exposed, and 133 controls, of whom 7 were exposed.

Ignoring biomass fuel exposure for the moment, perform an overall test for the association between age and tuberculosis and report a p-value. Now perform a test for trend to evaluate the evidence for increasing risk of tuberculosis with age.

Estimate the Odds Ratio for tuberculosis associated with indoor pollution for each age stratum separately. Qualitatively describe your assessment of how age modifies the effect of biomass fuel exposure on tuberculosis.

Question 11.3

The data set *oesoph*, found by going to http://www.crcpress.com/e_products/downloads/, contains more detailed information on alcohol consumption; see Question 7.3. Using the full data on four levels of alcohol consumption, perform an overall test for the association between alcohol consumption and incidence of esophageal cancer. Also assess the evidence for increasing risk over the four increasing levels of alcohol consumption, with an appropriate choice of ordered code, k, say, for each of the consumption groups. Make two plots with (1) $\hat{p}_k = \hat{P}$ (case|alcohol level k) and (2) OR_k (Odds Ratio for kth level vs. baseline level of 0 to 39 g/day alcohol consumption) on the Y-axis, both against k on the X-axis. Interpret the results of the two tests in light of these plots.

Question 11.4

Using the data set, *titanic*, found by going to http://www.crcpress.com/e_products/downloads/, examine the existence of a trend in the risk of dying as a passenger's ticket class changes from 1st to 2nd to 3rd. Plot the Relative Risk of dying for each ticket class using the 1st class passengers as the reference group, and plot the three Relative Risks against the ticket class.

4. Investigate the relationship between alcohol consumption and incidence of esophageal cancer. First, treat alcohol consumption as a categorical variable with four categories (0 to 39 g/day, 40 to 79 g/day, 80 to 119 g/day, and 120 g/day), by using indicator variables for the various categories (select 0 to 39 g/day as the reference group); second, treat alcohol consumption as an ordered variable by appropriately coding the four categories of consumption. Compare the two analyses and discuss whether an increasing trend in risk, as alcohol consumption increases, adequately fits the pattern of risks for the four categories. Compare your interpretations with those from Question 11.3.

Question 14.3

Referring again to the data set *titanic*, found by going to http://www.crcpress.com /e_products/downloads/, expand the analysis of Question 13.2 by fitting a logistic regression model that includes both ticket class (use indicator variables) and fare (interpret the roles of both risk factors, and discuss confounding issues by comparison to your single variable models of Question 13.2). Repeat this analysis, but now adding the passenger's sex to the model; again comment on the issue of confounding. Now add the variable for place of embarkment and revise your interpretations as necessary.

Question 14.4

Referring again to the data set, *titanic*, found by going to http://www.crcpress.com /e_products/downloads/, expand the analysis of Question 14.3 by fitting logistic regression models that include both ticket class (examine both indicator variables and an ordered scale) and sex, allowing for the possibility of multiplicative interaction between the two variables. What is your assessment and interpretation of interaction? In the model with interaction, does the ordered scale for ticket class allow for an adequate fit as compared with using indicator variables? Compare with Question 13.2 and discuss.

Question 14.5

The data set *breastcancer*, found at http://www.crcpress.com/e_products/downloads/, contains data from a population-based study of 3303 women in Nashville with prior benign breast biopsies (Dupont and Page, 1985). Information on these previous biopsies included whether the biopsied lesion was proliferative or not, as well as the family history of breast cancer. Women with prior proliferative breast lesions were further classified by whether the lesions also displayed atypical hyperplasia. The women were then followed for around 17 years for subsequent incidence of breast cancer. Using logistic regression, examine the roles of proliferative breast disease, with and without atypical hyperplasia, and family history on the incidence of breast cancer. Consider interactive effects if and when appropriate.