

Categorical Data Analysis

Homework I

24/03/2021

1. In a crossover trial comparing a new drug to a standard, p denotes the probability that the new one is judged better. It is designed to estimate p and test $H_0 : p = 0.5$ versus $H_a : p \neq 0.5$. In 20 independent observations, the new drug is better each time.
 - (a) Write out and plot the likelihood function. Is it close to the quadratic shape that large-sample normal approximations utilize?
 - (b) Give the ML estimate of p . Conduct a Wald test and construct a 95% Wald confidence interval for p . Are these sensible?
 - (c) Conduct a score test, reporting the p -value. Construct a 95% score confidence interval. Interpret.
 - (d) Conduct a likelihood-ratio test and construct a likelihood-based 95% confidence interval. Interpret.
 - (e) Construct an exact binomial test. Interpret.
2. Suppose $Y \sim \text{Binom}(n, p)$. We assign a Beta prior distribution, $\text{Beta}(\alpha, \beta)$, to the binomial parameter p .
 - (a) Write out the Bayes estimator of p .
 - (b) Show that the ML estimator is a limit of Bayes estimators, for a certain sequence of Beta prior parameter values.
 - (c) Find a prior density such that the Bayes estimator coincides with the ML estimator. Is this prior proper?
 - (d) Recall the example of estimating the proportion of vegetarians, that is, each of $n = 25$ students has been asked whether he or she was a vegetarian and $y = 0$ answered “yes”. With the Jeffreys’ prior, find the posterior mean, posterior 95% equal-tail interval and 95% highest posterior density interval of p .
3. A research study estimated that under a certain condition, the probability that a subject would be referred for heart catheterization was 0.906 for whites and 0.847 for blacks.

- (a) A press release about the study stated that the odds of referral for cardiac catheterization for blacks are around 60% of the odds for whites. Explain how they obtained 60%.
- (b) An associated Press story later described the study and said “Doctors were only 60% as likely to order cardiac catheterization for blacks as for whites.” Explain what is wrong with this interpretation. Give the correct percentage for this interpretation.
4. The following table refers to applicants to graduate school at the University of California at Berkeley, for fall 1973. It presents admission decisions by gender of applicant for the six largest graduate departments. Denote the three variables by A = whether admitted, G = gender, and D = department. Find the sample AG conditional odds ratios and the marginal odds ratio. Interpret, and explain why they give such different indications of the AG association.

Department	Whether Admitted			
	Male		Female	
	Yes	No	Yes	No
A	512	313	89	19
B	353	207	17	8
C	120	205	202	391
D	138	279	131	244
E	53	138	94	299
F	22	351	24	317
Total	1198	1493	557	1278

Source: Data from Freedman et al. (1978, p.14). See also P. Bickel et al., *Science* **187**: 398–403 (1975).

5. For a $2 \times 2 \times 2$ table, show that homogeneous association is a symmetric property.
6. For 2×2 tables, with multinomial sampling, the asymptotic variance of $\log \hat{\theta}$ is $\sum_i \sum_j 1/(np_{ij})$. Yule's Q is defined as

$$Q = \frac{p_{11}p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}}.$$

Use the asymptotic variance of $\log \hat{\theta}$ to work out the asymptotic variance of Q .

7. Refer to the table below.

- (a) Using X^2 and G^2 , test the hypothesis of independence between party identification and race. Report the p -values and interpret.
- (b) Use **standardized residuals** to describe the evidence of association.
- (c) **Partition chi-square into components** regarding the choice between Democrat and Independent and between these two combined and Republican. Interpret.

	Party Identification		
Race	Democrat	Independent	Republican
Black	192	75	8
White	459	586	471