STAT 854. Biometrics: Methods in Biomedical Studies (Spring 2023)

Homework 1: due on Monday Jan 23 2023 by 5:00pm via the course canvas page

Problem 1. What percent of lung cancer deaths among Canadian women may be attributed to smoking?

- Clearly state how you arrived at your answer.
- To answer this question you will need to find published information relevant to the quantity. Give the reference citation in your answer. One relevant information resource is the web site of *Canadian Cancer Society*: www.cancer.ca. Feel free to use other available information.
- Please be reminded that a confidence interval, say, at the level of 95%, is usually more welcomed than a point estimate in practice.

Problem 2. Assume that the four frequency counts in a 2×2 contingency tables, denoted by $N_{11}, N_{12}, N_{21}, N_{22}$, are independent and following the Poisson distributions with the parameters μ_{ij} for $i = 1, 2$ and $j = 1, 2$, respectively. Recall that the relative risk $RR = \frac{\mu_{11}/\mu_{1.}}{\mu_{21}/\mu_{2.}}$ and the odds ratio $OR = \frac{\mu_{11}/\mu_{12}}{\mu_{21}/\mu_{22}}$.

- What is the joint distribution of the counts $N_{11}, N_{12}, N_{21}, N_{22}$ from a cohort study with size n predetermined? [This sampling scheme is called by "multinomial sampling".] Give the MLE of RR and OR with the contingency table.
- What is the joint distribution of N_{11} and N_{21} from a case-control study with the sizes of the case group and the control group predetermined (i.e. the row subtotals $N_{1.} = n_{1.}$ and $N_{2.} = n_{2.}$)? [This sampling scheme is called by "purposive sampling".] Give the MLE of RR and OR with the contingency table.

Problem 3. Assume the hazard rate of a disease is constant across all subjects in a population for some time period. Joan Hu was not sure about why the following gives an approximate 95% confidence interval for the constant hazard rate,

$$\frac{c}{PT} \pm 1.96 \frac{\sqrt{c}}{PT},$$

where c=num of observed disease cases in PT person-time, the total time at risk of the group in the population where the cases were observed. Help her understand the approach.

State explicitly what assumptions are needed in your explanation.

Problem 4. Survival from age 35 for continuing cigarette smokers and lifelong non-smokers is presented in Figure 3 of Doll et al. (2004).[1] These data come from 50 years of observations on male British physicians recruited in 1951. Use these data to summarize the effect of smoking on mortality. Any relevant comments are welcomed.

---

[1] Doll R, Peto R et al. "Mortality in relation to smoking: 50 years observations on male British doctors". British Medical Journal 2004 (June 26); 328: 1519-1528.