

TA's notes

- Whenever possible, parts of a question will be graded conditionally on how you answered the preceding part(s). Be consistent throughout your problem solving.
- You may feel free to discuss homework with others. However, You should only turn in your own individual work (Please refer to the university policies on academic honesty and student conduct).

Problem 2.1 [8 points]**a. [4 points]**

In this case, there is no strong evidence of causal linkage between these two variables. You should consider both variables as response variables, with a joint distribution, caused by some other explanatory variables. Calculating the odds ratio, θ , will be the best way to describe the association. Yet credits will be also given if you assign the explanatory and response distinction according to the table layout: e.g., Explanatory variable: Gun Registration (2 levels - favor, oppose). Response variable: Death Penalty (2 levels - favor, oppose).

b. [4 points]

The sample odds ratio is $\frac{784 \cdot 66}{236 \cdot 311} = .705$

The odds for favoring death penalty is .705 times lower for people who favor gun registration than those who oppose (gun registration). People favoring gun registration are .705 times as likely as those opposing gun registration to favor death penalty.

You may use either of the following two measures for the association, if you assigned the explanatory and response distinction:

1. Difference of Proportion

Of $(784+236)=1020$ people who favor gun registration, there were 784 also favoring death penalty, a proportion of .769; Of $(311+66)=377$ people who oppose gun registration, there were 311 favoring death penalty, a proportion of .825. The sample difference of proportions is -.056. That is, there is a slightly bigger proportion (5.6%) of people who favor death penalty among the ones who oppose gun registration than the ones who are in favor (of gun registration).

2. The relative risk is $.769/.825=.932$

The proportion of favoring death penalty is 93.2% smaller for those who who favor gun registration than for those who oppose (gun registration).

Problem 2.2 [3 points]

“The odds ratio between treatment (A, B) and cure (yes, no) is 1.5.”

- The odds for treatment A to cure ($\pi_{yes|A}/\pi_{no|A}$) is 1.5 times greater than the odds for treatment B to cure ($\pi_{yes|B}/\pi_{no|B}$), i.e., The odds of cure is 1.5 times greater for treatment A than for treatment B.
- We do not have information regarding what the probability of treatment A to cure ($\pi_{yes|A}$) is, nor do we know the probability of treatment B to cure ($\pi_{yes|B}$). Thus, the interpretation of “the probability of cure is 1.5 times higher for treatment A than for treatment B” is incorrect. It is an interpretation of the relative risk ($\pi_{yes|A}/\pi_{yes|B}$).

For any cumulative period of daily average number of cigarettes, the control group systematically has the same or more patients than the lung cancer group. That is, the lung cancer group is stochastically higher than the control group with respect to their distributions on smoking of cigarettes.

b. [4 points]

Daily Average No. of Cigarettes	Lung Cancer Group	Control Group
< 5	62	190
>= 5	1295	1167

The odds ratio is $\frac{62 \cdot 1167}{190 \cdot 1295} = .294$. That is, the odds for male patients to develop lung cancer is .294 times lower for those who on average smoked fewer than 5 cigarettes per day over a ten-year period than those who on average smoked at least 5.

c. [4 points]

No, we cannot estimate the difference in the proportions who got lung cancer between those who smoked fewer than 5 cigarettes per day and those who smoked at least 5 per day. This is a retrospective study, so we do not know the population prevalence of lung cancer. These data may have been obtained from “samples of convenience” (whatever hospitals in several English cities had). We do not have information regarding how and when the data were collected. You may calculate the quantity according to the previous table:

$$\frac{62}{62 + 190} - \frac{1295}{1295 + 1167} = .246 - .526 = -.280$$

However, this quantity does not reflect the difference of the population proportions, and thus is not an estimate for the difference.

Problem 2.8 [8 points]

a. [3 points]

Within row i , the odds that a subject would be in the lung cancer group instead of the control group is defined to be:

$$\Omega_i = \frac{\pi_{cancer|i}}{\pi_{control|i}}$$

Therefore, we can add the sample log odds for each level of smoking to the table:

Daily Average Number of Cigarettes	Lung Cancer Group	Control Group	Log(Odds)
0	7	61	$\log(\Omega_{none}) = \log(7/61) = -2.165$
< 5	55	129	$\log(\Omega_{<5}) = \log(55/129) = -.852$
5-14	489	570	$\log(\Omega_{5-14}) = \log(489/570) = -.153$
15-24	475	431	$\log(\Omega_{15-24}) = \log(475/431) = .097$
25-49	293	154	$\log(\Omega_{25-49}) = \log(293/154) = .643$
50+	38	12	$\log(\Omega_{50+}) = \log(38/12) = 1.15$

As level of smoking increases, there is an increasing trend in the sample log odds. The odds that male smokers had lung cancer (rather than other diseases) increases as their daily average number of cigarettes increases. In addition to the increasing trend, there is a sign change at level 3. It indicates that there are more control patients than lung cancer patients among those who smoked fewer than 15 cigarettes; whereas there are more lung cancer patients than control patients among those who smoked at least 15 cigarettes per day.

b. [3 points]

The 5 log local odds ratios for each pair of adjacent level of smoking are as follows.

$$\begin{aligned} \log(\theta_{12}) &= \log(7 * 129 / 61 * 55) = -1.312 \\ \log(\theta_{23}) &= \log(55 * 570 / 489 * 129) = -.699 \\ \log(\theta_{34}) &= \log(489 * 431 / (570 * 475)) = -.250 \\ \log(\theta_{45}) &= \log(475 * 154 / 431 * 293) = -.546 \\ \log(\theta_{56}) &= \log(293 * 12 / 154 * 38) = -.509 \end{aligned}$$

The 5 log local odds ratios were all negative, i.e., people who had the higher level of smoking always had the greater odds to be in the lung cancer group. It is more so for the comparison between people who never smoked and those who smoke fewer than 5 cigarettes per day. Further, we can find that the log local odds ratio is the corresponding difference between each pair of adjacent log odds (cf. part(a.)). The nature of the association between level of smoking and lung cancer is not linear. Should the nature of the association be linear, we would have found the log local odds identical (consistent with the slope).

c. [2 points]

We know the log local odds ratio for each pair of adjacent levels of smoking can be expressed as:

Given $\log(odds_i) = \alpha + \beta i$

$$\begin{aligned} \log \theta_{i,i+1} &= \log\left(\frac{\Omega_i}{\Omega_{i+1}}\right) \\ &= \log \Omega_i - \log \Omega_{i+1} \\ &= \alpha + \beta i - (\alpha + \beta(i + 1)) \\ &= -\beta \end{aligned}$$

Hence, if the log odds of lung cancer was linearly related to the level of smoking, all log local odds ratios would be identical as a constant value $(-\beta)$, i.e., all local odds ratios are identical ($\exp^{-\beta}$). You can also find this relationship by graph.