STA 138: Fall 2006
Homework 2 Solutions

Problems: 2.2, 2.4, 2.10, 2.12, 2.20, 2.22, 2.24, 2.36

2.2 A diagnostic test has sensitivity=specificity=0.80. Find the odds ratio between true disease status and the diagnostic test result.

Solution:

Let D denote disease and $\bar{D}$ no disease. Similarly let, T denote positive test and $\bar{T}$ negative test. The odds ratio is

$$\theta = \frac{P(T|D)P(\bar{T}|\bar{D})}{P(T|\bar{D})P(\bar{T}|D)}.$$

Now, sensitivity= $P(T|D) = 0.80$ and specificity= $P(\bar{T}|\bar{D}) = 0.80$. Hence, $\theta = (0.8 \times 0.8)/(0.2 \times 0.2) = 16$.

2.4 Consider the following two studies reported in the *New York Times*.

a. A British study reported (Dec. 3, 1998) that of smokers who get lung cancer, "women were 1.7 times more vulnerable than men to get small-cell lung cancer." Is 1.7 the odds ratio or the relative risk?

b. A National Cancer Institute study about tamoxifen and breast cancer reported (Apr. 7, 1998) that the women taking the drug were 45% less likely to experience invasive breast cancer than were women taking placebo. Find the relative risk for (i) those taking the drug compared to those taking placebo, and (ii) those taking placebo compared to those taking drug.

Solution:

a. Let π_1 denote the probability a woman is vulnerable to get small-cell lung cancer, and π_2 the probability for men. Then the report says $\pi_1 = 1.7\pi_2$. So $1.7 = \frac{\pi_1}{\pi_2}$ is the relative risk.

b. Let π_1 denote the probability a women taking the drug experiences invasive breast cancer, and π_2 the probability for women taking placebo. Then the report says $\pi_1 = (1 - 45\%)\pi_2 = 0.55\pi_2$. So the relative risk for (i) those taking the drug compared to those taking placebo is

$$\frac{\pi_1}{\pi_2} = 0.55,$$

and the relative risk for (ii) those taking placebo compared to those taking drug is

$$\frac{\pi_2}{\pi_1} = 1/0.55 = 1.8182.$$

2.10 A research study estimated that under a certain condition, the probabiltiy 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 60% of the odds for whites. Explain how they obtained 60% (more accurately, 57%).

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.

Solution:

a. The odds ratio is calculated as $\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{(0.847)(1-0.906)}{(0.906)(1-0.847)} = 0.57$.

b. The percentage 0.60 reported is for the odds ratio but the interpretation is for the relative risk. The correct relative risk is $RR = \pi_{11}/\pi_{12} = 0.847/0.906 = 0.935$.

2.12 Table 2.10 refers to applicants to graduate school at the University of California at Berkeley, for fall 1973. It presents admissions 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.

Solution:

The sample AG conditional odds ratios are:

$$\theta_{AG(A)} = 0.3492;\ \theta_{AG(B)} = 0.8025;\ \theta_{AG(C)} = 1.1331;$$

$$\theta_{AG(D)} = 0.9213;\ \theta_{AG(E)} = 1.2216;\ \theta_{AG(F)} = 0.8279.$$

The marginal odds ratio is: $\theta_{AG} = 1.8411$.

Hence in department A the sample odds for male being admitted was 0.3492 times the sample odds for a female. In department B, C, D, E, and F, they were 0.8025, 1.1331, 0.9213, 1.2216, and 0.8279 respectively. However for the marginal odds ratio, the sample odds for a male being admitted were 84% higher than for a female. Within each department, these odds were smaller for males. The different indications of the AG association after controlling for departments illustrate Simpson's paradox. The reason is that the association between department and gender is very strong. For example, the marginal odds ratio for department A and E with variable gender is $(825 \times 393)/(191 \times 108) = 15.7177$.

2.20 Table 2.13 is from an early study on the death penalty in Florida. Analyze these data and show that Simpson's paradox occurs.

Solution:

Let X, Y, Z denote Defendant's Race (white, black); Death Penalty (yes, no); and Victim's Race (white, black) respectively. Then for white victim's race, the relative risk of death penalty for white defendant to black is $(19/151)/(11/63) = 0.7207$, and the odds ratio is $\theta_{XY(Z)} = (19 \times 52)/(11 \times 132) = 0.6804$.

For black victims, both the relative risk and odds ratio are 0 since the proportion of white defendants receiving the death penalty is 0. However for marginal odds ratio, $\theta_{XY} = (19 \times 149)/(17 \times 141) = 1.1811$, which is greater than the odds within each victim's race. The result shows that Simpson's paradox occurs for these data.

2.22 Binomial parameters for two groups are graphed with π_1 on the horizontal axis and π_2 on the vertical axis. Plot the locus of points for a 2×2 table having

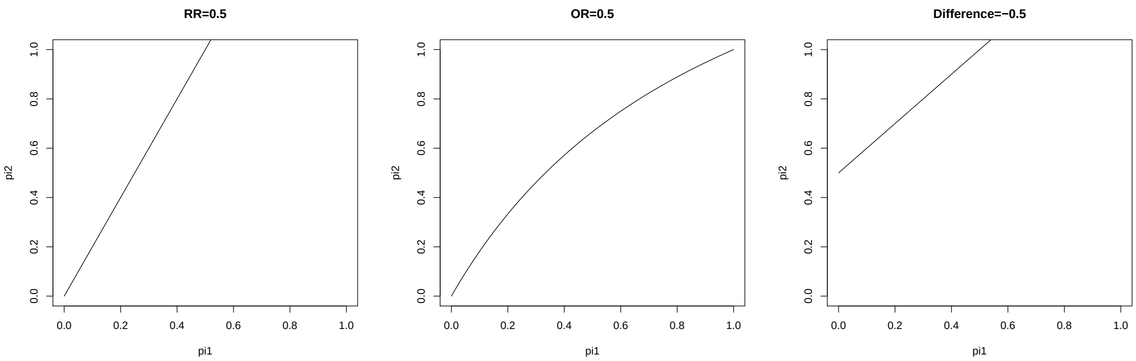
a. relative risk = 0.5,

b. odds ratio = 0.5, and

c. differnce of proportions = -0.5.

Solution:

a. relative risk $= 0.5 \Rightarrow \pi_2 = 2\pi_1$.
b. odds ratio $= 0.5 \Rightarrow \pi_2 = \pi_1/(0.5 + 0.5\pi_1)$.
c. differnce of proportions = -0.5 $\Rightarrow \pi_2 = \pi_1 + 0.5$.



2.24 For a 2×2 table of counts $\{n_{ij}\}$ show that the odds ratio is invariant to

a. Interchanging rows with columns, and
b. Multiplication of cell counts within rows or columns by $c \neq 0$.

Show that the difference of proportions and the relative risk do not have these properties.

Solution:

First, consider the odds ratio.

a. Interchanging rows with columns: $n_{ij} = n_{ji}$. The odds ratio becomes $\theta = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{n_{22}n_{11}}{n_{21}n_{12}}$
b. Multiplication of cell counts within rows or within columns by c: the odds ratio becomes $\theta = \frac{cn_{11}n_{22}}{cn_{12}n_{21}} = \frac{n_{11}cn_{22}}{n_{12}n_{c21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$. Hence, the odds ratio is invariant.

However, for the difference of proportions:

a. $\frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{22}}{n_{21}+n_{22}} \neq \frac{n_{22}}{n_{21}+n_{22}} - \frac{n_{11}}{n_{11}+n_{12}}$
b. $\frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{22}}{n_{21}+n_{22}} \neq \frac{cn_{11}}{cn_{11}+n_{12}} - \frac{n_{22}}{cn_{21}+n_{22}} \neq \frac{n_{11}}{n_{11}+cn_{12}} - \frac{cn_{22}}{n_{21}+cn_{22}}$

And for relative risk:

a. $\frac{\frac{n_{11}}{n_{11}+n_{12}}}{\frac{n_{22}}{n_{21}+n_{22}}} \neq \frac{\frac{n_{22}}{n_{21}+n_{22}}}{\frac{n_{11}}{n_{11}+n_{12}}}$

b. $\frac{\frac{n_{11}}{n_{11}+n_{12}}}{\frac{n_{22}}{n_{21}+n_{22}}} \neq \frac{\frac{cn_{11}}{cn_{11}+n_{12}}}{\frac{n_{22}}{cn_{21}+n_{22}}} \neq \frac{\frac{n_{11}}{n_{11}+cn_{12}}}{\frac{cn_{22}}{n_{21}+cn_{22}}}$

2.36 For 2×2 tables, Yule (1900, 1912) introduced

$$Q = \frac{\pi_{11}\pi_{22} - \pi_{12}\pi_{21}}{\pi_{11}\pi_{22} + \pi_{12}\pi_{21}},$$

which he labeled Q in honor of the Belgian statistician Quetelet. It is now called Yule's Q.

a. Show that for 2×2 tables, Goodman and Kruskal's $\gamma = Q$.

b. Show that Q falls bewteen -1 and 1.

c. State the conditions under which $Q = -1$ or $Q = 1$.

d. Show that Q relates to the odds ratio by $Q = (\theta - 1)/(\theta + 1)$, a monotone transformation of θ from the $[0, \infty]$ scale onto the $[-1, 1]$ scale.

Solution:

a. Goodman and Kruskal's γ is

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d},$$

where $\Pi_c = 2\sum_i \sum_j \pi_{ij}(\sum_{h>i} \sum_{k>j} \pi_{hk})$ and $\Pi_d = 2\sum_i \sum_j \pi_{ij}(\sum_{h>i} \sum_{k<j} \pi_{hk})$. For a 2×2 table, $\Pi_c = 2\pi_{11}\pi_{22}$ and $\Pi_d = 2\pi_{12}\pi_{21}$. So,

$$\gamma = \frac{2\pi_{11}\pi_{22} - 2\pi_{12}\pi_{21}}{2\pi_{11}\pi_{22} + 2\pi_{12}\pi_{21}} = Q.$$

b. Note that all $\pi_{ij} \geq 0$. The maximum of γ occurs when $\pi_{12}\pi_{21} = 0$, since this value makes the numerator largest. When $\pi_{12}\pi_{21} = 0$, $\gamma = \pi_{11}\pi_{22}/\pi_{11}\pi_{22} = 1$. Now, the minimum of γ occurs when $\pi_{11}\pi_{22} = 0$, or when $\gamma = -\pi_{12}\pi_{21}/\pi_{12}\pi_{21} = -1$. Hence the range of γ is $[-1, 1]$.

c. As described above, $Q = 1$ when at least one of π_{12} or π_{21} is zero and hence at least one of π_{11} or π_{22} is one. And $Q = -1$ when at least one of π_{11} or π_{22} is zero and hence at least one of π_{12} or π_{21} is one. In other words, when there are no counts in one of the cells Q is at a maximum or minimum. You can see this in the following algebra:

$$Q = 1 \Rightarrow \pi_{11}\pi_{22} - \pi_{12}\pi_{21} = \pi_{11}\pi_{22} + \pi_{12}\pi_{21} \Rightarrow -\pi_{12}\pi_{21} = \pi_{12}\pi_{21}$$

Which is true only if one of π_{12} or π_{21} is zero. A similar argument holds for $Q = -1$.

d. Consider the following monotone transformation of θ from the $[0, \infty]$ scale onto the $[-1, 1]$ scale:

$$\begin{aligned} f(\theta) &= \frac{\theta - 1}{\theta + 1} \\ &= \frac{\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} - 1}{\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} + 1} \\ &= \frac{\frac{\pi_{11}\pi_{22} - \pi_{12}\pi_{21}}{\pi_{12}\pi_{21}}}{\frac{\pi_{11}\pi_{22} + \pi_{12}\pi_{21}}{\pi_{12}\pi_{21}}} \\ &= \frac{\pi_{11}\pi_{22} - \pi_{12}\pi_{21}}{\pi_{11}\pi_{22} + \pi_{12}\pi_{21}} \\ &= Q. \end{aligned}$$