13 September 2007

Estimating an Odds Ratio

Eric Rasmusen, erasmuse@Indiana.edu, http://www.rasmusen.org/x/2007/oddsratio.pdf

Let $n(radon)$ denote the probability someone was exposed to radon and $n(unexposed) = 1 - n(radon)$ be the probability he was not. Let $P(cancer|radon)$ be the probability someone gets cancer after having been exposed to radon and $P(healthy|radon)$ be the probability he does not get cancer, and let $P(cancer|unexposed)$ and $P(healthy|unexposed)$ be the equivalents for the unexposed. Finally, let θ and γ be the probability that cancer and helathy observations are sampled. Then the expected proportions in the sample are as shown in this table:

	Radon	Unexposed
Cancer	$\theta P(cancer|radon) \cdot n(radon)$	$\theta P(cancer|unexposed) \cdot n(unexposed)$
Healthy	$\gamma P(healthy|radon) \cdot n(radon)$	$\gamma P(healthy|unexposed) \cdot n(unexposed)$

The odds ratio is the ratio of the odds of getting cancer when exposed tor adon to the odds of getting cancer when unexposed:

$$\text{Odds Ratio} = \frac{\theta P(cancer|radon) \cdot n(radon)/[\gamma P(healthy|radon) \cdot n(radon)]}{\theta P(cancer|unexposed) \cdot n(unexposed)/[\gamma P(healthy|unexposed) \cdot \cdot n(unexposed)]}, \tag{1}$$

which can be simplified thus:

$$\text{Odds Ratio} = \frac{P(cancer|radon)/P(healthy|radon)}{P(cancer|unexposed)/P(healthy|unexposed)} \tag{2}$$

The sampling probabilities γ and θ have cancelled out. Thus, even if we sample cancer observations more heavily than healthy observations, we still will get the same expected value of the odds ratio.

This is true regardless of the functional form of the distribution of the disturbances. They can be normal, or logistic, or uniform. But it is especially noteworthy for the logistic distribution, because for that distribution, the most important parameter is β, which is exactly the odds ratio. Thus, the value computed using equation (1) is an unbiased estimate of β. If the disturbances had a normal distribution, on the other hand, we could still use equation (1) as an unbiased estimate of the odds ratio, but it would be a complicated function of the β and σ parameters and we could not recover parameter estimates from our observed values.

The odds ratio is not useful in itself. It is hard to understand. But it is an approximation of something more interesting: the relative risk. The relative risk is the ratio of the probability of getting cancer if radon-exposed to the probability of getting cancer if unexposed, and if cancer is rare whether a person is exposed or not ($P(healthy|radon) \approx 1 \approx P(healthy|unexposed)$) then the odds ratio is approximately equal to the relative risk:

$$\text{Relative Risk} = \frac{P(cancer|radon)}{P(cancer|unexposed)} \approx \frac{P(cancer|radon)/P(healthy|radon)}{P(cancer|unexposed)/P(healthy|unexposed)} = \text{Odds Ratio} \tag{3}$$