

(1) アリゾナ州に居住するピマ・インディアンには肥満の人が多いため、その原因を調べるため、ピマ・インディアンの住む村を訪問し、住民の成人からランダムサンプルした 150 人について簡単な健診と質問紙調査を行った（→ 一時点でアウトカムと要因の両方を調べるので断面研究になる）。実測した身長と体重から計算した BMI が 30 以上を肥満とし、食事摂取頻度調査でわかった、毎日一度はハンバーガーを食べる人をハンバーガー多食者として、そのクロス集計をした結果が、下表のようになったとする。

	肥満	肥満でない
ハンバーガー多食者	60	20
ハンバーガー多食でない	20	50

ここで、まず見るのは、実際に肥満がどれくらい多いのかを知るため、有病割合を計算してみることである。次に、ハンバーガー多食が肥満に対して効果があるかどうかを調べるため、効果指標としてオッズ比を計算する。有病割合とオッズ比を計算してみよう。

(2) ある工場で爆発事故があり、化学物質が周辺に飛散して、周辺住民がそれに曝露してしまった。どうい影響がでるのかわからないので、工場周辺の住民を曝露群、十分に離れたところに住んでいるけれども生活水準や人口構成や食習慣や職業を含む社会経済的背景に違いがない住民を非曝露群として、それぞれ 10000 人をランダムサンプルし、5 年間の追跡調査によって疾病発生状況を観察したところ（→ コホート研究になる）、曝露群では白血病の患者が 4 年目に 10 人、5 年目に 10 人発生し、非曝露群では 5 年目に 2 人だけ発生したとする。このとき、事故による化学物質への曝露が白血病発生に影響があったかどうかを調べるには、絶対的な効果をみるにはリスク差または罹患率差を計算し、相対的な効果をみるにはリスク比または罹患率比を計算する。計算してみよう。

(3) ある病院に急な嘔吐・腹痛・下痢を訴える患者が何人も来院したので、食中毒を疑い、それらの患者全員を症例群、そうした症状ではなく外傷等で来院した人を対照群として、最近 3 日間に何を食べたか等についての聞き取り調査を行った（→ 症例対照研究になる）。その結果、地元の特産品として何軒かの店で売られている食品 A を食べた人が、症例群 50 人中 45 人、対照群 100 人中には 10 人いた。食品 A 摂取の、この症状発症への効果をみるには、オッズ比を計算する（考えてみればわかると思うが、分母がわからないので有病割合は計算できないし、リスクも罹患率もわからないので、計算できるのはオッズ比だけである）。計算してみよう。

(1) の解答

断面研究で住民の成人全体からランダムサンプルした 150 人を調査して肥満者が 60+20=80 人だったので、有病割合は 80/150 より約 53% である。

オッズ比は $(60 \cdot 50) / (20 \cdot 20) = 7.5$ なので、ハンバーガーを多食すると 7.5 倍肥満しやすくなると言える。統計ソフトを使えば、オッズ比が 1 と差がない（つまり要因の有無と疾病の有無が無関係）という帰無仮説の検定や、オッズ比の 95% 信頼区間の推定もできる。<http://www.r-project.org> から入手できるフリーソフトである R を使えば、fmsb パッケージをインストールしてあれば（していなければ `install.packages("fmsb")` としてネットワークからインストールする）、

`library(fmsb); oddsratio(60, 20, 20, 50)` と打つだけで

オッズ比 : 7.5, 95% 信頼区間 [3.63, 15.5], p 値 = 1.45×10^{-8}

であることがわかる。

または、オッズ比の推定方法が最尤法であるため 7.5 にはならないがパッケージがなくても `fisher.test(matrix(c(60, 20, 20, 50), 2))` と打てば

オッズ比: 7.38, 95% 信頼区間 [3.43, 16.6], p 値 = 1.43×10^{-8}

という結果が得られる。

(2) の解答

曝露群の罹患率は $20 / (10 \times 3 + 10 \times 4 + 9980 \times 5) = 0.0004$ (/ 年), 5 年間のリスクは $20 / 10000 = 0.002$

非曝露群の罹患率は $2 / (2 \times 4 + 9998 \times 5) = 0.00004$ (/ 年), 5 年間のリスクは $2 / 10000 = 0.0002$

罹患率比は $0.0004 / 0.00004 = 10$, 5 年間のリスク比は $0.002 / 0.0002 = 10$

罹患率差は 0.00036 (/ 年), 5 年間のリスク差は 0.0018

95% 信頼区間や p 値を求めるには、R と fmsb パッケージを使って、

`library(fmsb)`

`rateratio(20, 2, 10 * 3 + 10 * 4 + 9980 * 5, 2 * 4 + 9998 * 5)`

`ratedifference(20, 2, 10 * 3 + 10 * 4 + 9980 * 5, 2 * 4 + 9998 * 5)`

`riskratio(20, 2, 10000, 10000)`

`riskdifference(20, 2, 10000, 10000)`

とすれば良い（結果表示は省略する）。罹患率比とリスク比は概ね同じ値になるが、罹患率が得られている場合は罹患率比を採用するのが普通である。

(3) の解答

症状の有無と食品 A 摂取の有無でクロス集計表を作ると

	症例群	対照群
食品 A 摂取あり	45	10
食品 A 摂取なし	5	90

となるので、オッズ比は、 $(45/5) / (10/90) = (45 \times 90) / (5 \times 10) = 81$

症例群は対照群に比べ、食品 A を食べた確率が 81 倍ということの意味するので、ほぼ間違いなく、食品 A の摂取がこの症状の主要な原因であると推論できる。

95% 信頼区間や p 値を求めるには、断面研究の場合と同じく R と fmsb を使えば

`library(fmsb); oddsratio(45, 5, 10, 90)` とすれば

オッズ比 : 81, 95% 信頼区間 [26, 252], p 値 $< 2.2 \times 10^{-16}$

であることがわかる。