

保健統計学 (11)

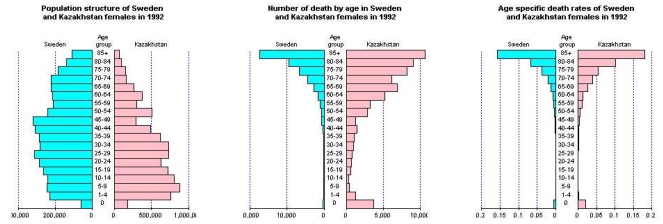
2016.6.7

- 交絡と回帰モデル
- 症例数とパワー計算：エビデンスベースヘルスケア特講 I のサンプルサイズの設計の話の裏返し
 - 一般にサンプルサイズが大きいほど検出力 (Statistical power) は大きくなる
 - 検出力は 1 から第二種の過誤 (本当は差があるのにそれを見逃す確率) を引いた値
- 多重検定：エビデンスベースヘルスケア特講で明日
 - 3群以上の比較をするとき、2群間の比較を繰り返すと全体としての第一種の過誤 (本当は差がないのに誤って差があると判定してしまう確率) が大きくなってしまいますので、個々の検定の有意水準あるいは有意確率を調整する (これを検定の多重性の調整という) 必要がある。多重比較ともいう
 - ボンフェローニの不等式に基づくボンフェローニの方法が最も単純。ホルムの方法や FDR が標準的。

交絡 (confounding) とは

- 原因→結果という因果関係を考えているとき、原因とも結果とも関連があって、しかも中間媒介因子ではない因子があるとき、これを交絡因子と呼ぶ。交絡因子の影響によって、注目している因果関係は歪められる。この歪みを交絡と呼ぶ
- 交絡要因の3条件 (例：肥満⇒高血圧に対する年齢)
 - 注目している要因 (肥満) ではない
 - 注目している要因 (肥満) と結果である健康影響 (高血圧) の両方と因果的に関連
 - 要因曝露の結果ではない (肥満が高齢をもたらすのではない)

交絡の例 (1)



- スウェーデンの女性とカザフスタンの女性の 1992 年の死亡率
- スウェーデンの方がカザフスタンより衛生水準・医療水準・生活水準が高い → 年齢別死亡率は全年齢層でスウェーデン < カザフスタン → 当然、平均寿命はスウェーデン > カザフスタン → が、粗死亡率はスウェーデン (10.5) > カザフスタン (6.3)
- 理由は以下 2 点
 - どちらの国でも高齢者の死亡率 >> 若者や成人の死亡率
 - 高齢者の割合はスウェーデン > カザフスタン
- 年齢構造が交絡因子となっている

交絡の例 (2)

- 年齢が交絡因子となっているために、喫煙者の方が非喫煙者よりも死亡率が低くなることもある (出典：『ロスマンの疫学』)

表 1-1 英国ウィッカムの女性における調査開始時の喫煙状況による 20 年間の死亡リスク

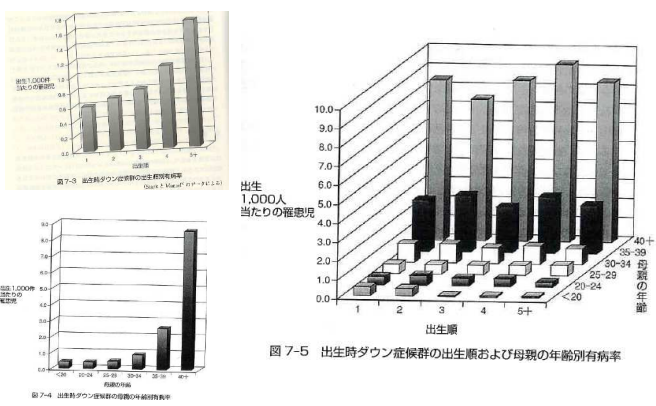
生存状況	喫煙者	非喫煙者	合計
死亡	139	230	369
生存	443	502	945
総数	582	732	1314
リスク (死亡/総数)	0.24	0.31	0.28

表 1-2 英国ウィッカムの女性における調査開始時の年齢別喫煙状況による 20 年間の死亡リスク

年齢 (歳)	生存状況	喫煙者	非喫煙者	総数
18-24	死亡	2	1	3
	生存	53	61	114
25-34	死亡リスク	0.04	0.02	0.03
	生存	121	152	273
35-44	死亡リスク	0.02	0.03	0.03
	生存	95	114	209
45-54	死亡	27	12	39
	生存	103	66	169
55-64	死亡リスク	0.21	0.15	0.19
	生存	64	40	104
65-74	死亡	29	101	130
	生存	7	28	35
75+	死亡リスク	0.81	0.78	0.79
	生存	13	64	77
	死亡	0	0	0
	生存	1.00	1.00	1.00

交絡の例 (3)

- ダウン症ハイリスクなのは出生順位が後の子か母の年齢が高い子か？ (出典：『ロスマンの疫学』)



交絡の制御

- デザインによる制御
 - 性別や年齢による交絡がありそうな場合は、それらで層別した層別無作為抽出をする。コホート研究なら曝露群と非曝露群を設定するときに年齢や性別をマッチングする、特定の年齢や性別に限定化して研究すると、それらの影響がデータに表れない
- 解析方法の工夫による制御
 - 層別解析：層別に比較する
 - 標準化：年齢調整死亡率とか
 - プール化：どの層でも同じ関連や違いがあると仮定して要約統計量を計算する (マンテルヘンツェルの要約オッズ比など)
 - 多変量解析：交絡因子も独立変数に含めたロジスティック回帰分析等

ロジスティック回帰分析の例

- エコポイントチェックデータ
 - <http://minato.sip21c.org/advanced-statistics/ecopx.txt>
 - <http://minato.sip21c.org/humeco/ecopoint.html>

エコポイント・チェック (Eco-Point Check)

Last updated on January 9, 2008 (08:03 14:40 (リンク先更新))

環境問題・経済問題の両面から注目されているエコポイント制度の普及が促進されています。環境問題の解決に貢献するだけでなく、経済効果も期待されています。エコポイント制度の普及に伴って、環境問題の解決に貢献するだけでなく、経済効果も期待されています。エコポイント制度の普及に伴って、環境問題の解決に貢献するだけでなく、経済効果も期待されています。

エコポイント・チェック (Eco-Point Check)

エコポイント・チェック (Eco-Point Check)

エコポイント・チェック (Eco-Point Check)

質問項目

AGE	年齢 (10歳階級)	0=10-19	1=20-29	2=30-39	3=40-49	4=50-59	5=60-69	6=70+
SEX	性別	0=M	1=F					
FAMSIZE	自分を含む同居人数							
Q01	新聞雑誌リサイクル	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q02	資源 (100%) トイレペーパー使用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q03	飲料容器リサイクル	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q04	買い物袋持参	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q05	冷暖房より着る服で調節	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q06	食料を期限切れで捨てない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q07	風呂は家族で回して入る (一人暮らしの場合はお湯を少くする)	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q08	風呂水を洗濯等に利用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q09	車のアイドリングストップ (車を持っていない人は「いつも」になる)	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q10	マイカーを上げて公共交通を利用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q11	太陽熱温水器を利用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q12	家電製品は省エネ型以外を買わない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q13	米のとぎ汁は煮かず有効利用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q14	油をふき取ってからゴミ捨	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q15	塩じ薬グッズ製品を買わない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q16	洗剤は含塩洗剤でない洗剤を利用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q17	洗剤 (石鹸を含む) をはかって適量使用	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q18	除草剤や殺虫剤を使わない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q19	車のバッテリーや電池を適正処理	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q20	トイレや風呂の強力洗剤を利用しない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q21	有機洗剤を利用しない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q22	有機農産物を選ぶ	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q23	地域の農産物を選ぶ	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q24	草履草起す	0=いつも	1=大体	2=時々	3=たまに	4=皆無		
Q25	煙草を吸わない	0=いつも	1=大体	2=時々	3=たまに	4=皆無		

ロジスティック回帰

- Q4 と Q10 の関係には年齢と性別の影響が両方ある (年齢と性別は Q4 と Q10 の関係を交絡している) → 年齢と性別と Q10 を独立変数、Q4 を従属変数としてロジスティック回帰分析 → 年齢と性別の影響を調整して、Q10 が TRUE の人 (いつも / 大体 / 時々 公共交通便利 = 公共交通利用者) は FALSE の人 (公共交通利用がたまに / 皆無 = マイカー族) に比べて、どれくらい Q4 が TRUE である (いつも / 大体 / 時々 買い物袋持参 = エコバッグ派である) 割合が高いかわかる
- Q4 が TRUE である割合を p とすると、 $\log(p/(1-p)) = \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{SEX} + \beta_3 \text{Q10}$
TRUE を 1, FALSE を 0 とすると、他の変数を調整した上で Q10 の効果は、
 - Q10 が TRUE の場合の Q4 が TRUE である割合を p_1 とすれば、 $\log(p_1/(1-p_1)) = \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{SEX} + \beta_3$
 - Q10 が FALSE の場合の Q4 が TRUE である割合を p_0 とすれば、 $\log(p_0/(1-p_0)) = \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{SEX}$
 - 辺々引けば $\log(p_1/(1-p_1)) - \log(p_0/(1-p_0)) = \beta_3$
- 結局右の左辺は、 $\log(p_1/(1-p_1)) / (p_0/(1-p_0))$ となる。これは対数オッズ比である。つまり、AGE と SEX の影響を調整した上で Q10 が TRUE である人は FALSE の人に比べて、 $\exp(\beta_3)$ 倍、Q4 が TRUE になりやすい
- 「性別や年齢の影響を調整した上で、公共交通利用者はマイカー族に比べて、エコバッグ派である可能性が $\exp(\beta_3)$ 倍大きい」ということである。 $\beta_0 \sim \beta_3$ は、一般化線型モデルの係数として推定される。
- R でのプログラム


```
x <- read.delim("http://minato.sip21c.org/advanced-statistics/ecopx.txt"); attach(x)
res <- glm(Q4~Q10+AGE+SEX, family=binomial(logit)); summary(res); exp(coef(res))
```