

医学情報処理演習第 10 回「クロス集計」課題の解答例

課題

東南アジアからオセアニアにかけての広い地域に、ビンロウジュの房状の実の種を、石灰を水で少し溶いたものと一緒にある種の葉で包んで噛む習慣がある。口の中が真っ赤になり、覚醒剤の様に体が熱くなるため、習慣性がある。betel-chewing と呼ばれている。betel-chewing は古くから口腔がんのリスクファクターとなることが指摘されているが、喫煙との交絡も指摘されており、これまでの研究結果も一貫性がなかった。パプアニューギニアの口腔がん患者群 143 人と対照群 477 人についての症例対照研究の結果、毎日 betel-chewing をする者が患者群で 118 人（うち、毎日喫煙する者が 107 人、しない者が 11 人）、対照群で 338 人（うち、毎日喫煙する者が 272 人、しない者が 66 人）いて、betel-chewing が毎日ではない者が患者群で 6 人（うち、毎日喫煙する者が 4 人、しない者が 2 人）、対照群で 42 人（うち、毎日喫煙する者が 22 人、しない者が 20 人）であった。

このデータから、betel-chewing と口腔がんには関連があるか、あるとしたらどの程度の関連か評価せよ。まず毎日喫煙する者としない者で別々に 2 つのクロス集計表を作って別々に分析してから、その結果から判断して、必要な場合には、喫煙するしないによらず共通した関連があるかどうか、あるとすればどの程度の関連か評価せよ^a。

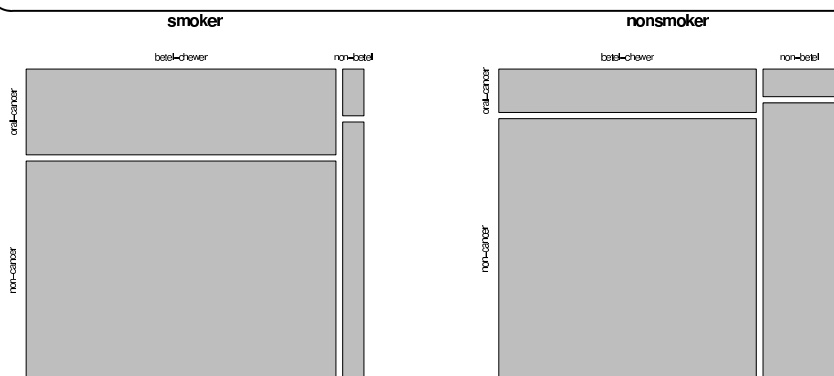
図を Powerpoint または Word に貼り付け、検定結果、それに対する考察、学籍番号、氏名もタイプして印刷し、署名して提出すること。

^a 出典：Thomas SJ et al. (2007) Betel quid not containing tobacco and oral cancer: A report on a case-control study in Papua New Guinea and a meta-analysis of current evidence. *Int. J. Cancer*, 120: 1318-1323. なお、原論文ではロジスティック回帰分析により性・年齢などの影響が調整されたオッズ比が計算されているが、本日の課題では他の影響は無視して分析する。

課題文の通り、素直に R のコードを書けば、以下のようになる。

kadai10-2007.R

```
smoker <- matrix(c(107,4,272,22),nr=2)
rownames(smoker) <- c("betel-chewer","non-betel")
colnames(smoker) <- c("oral-cancer","non-cancer")
nonsmoker <- matrix(c(11,2,66,20),nr=2)
rownames(nonsmoker) <- c("betel-chewer","non-betel")
colnames(nonsmoker) <- c("oral-cancer","non-cancer")
layout(t(1:2))
mosaicplot(smoker)
mosaicplot(nonsmoker)
fisher.test(smoker)
fisher.test(nonsmoker)
print(all <- array(c(smoker,nonsmoker),dim=c(2,2,2)))
library(vcd)
woolf_test(all)
mantelhaen.test(all)
assocstats(smoker)
assocstats(nonsmoker)
```



フィッシャーの直接確率を求めると、毎日喫煙する群 ($p=0.1794$) でも、しない群 ($p=0.7269$) でも、毎日 betel-chewing をするかどうかと口腔がん患者かどうか独立であるという帰無仮説は有意水準 5% で棄却されない。したがって、両者に関連があるとはいえない。オッズ比は喫煙群で 2.16 (95% 信頼区間: 0.71-8.83)、非喫煙群で 1.66 (95% 信頼区間: 0.32-16.7) と、点推定量は 1 より大きいですが、信頼区間の幅が広すぎるので、この結果だけでは何ともいえない。(層別で関連がなかったのでクロス表の併合には意味がなく、マンテル=ヘンツェルの結果は解釈しない)

属性相関係数は喫煙群で 0.07、非喫煙群で 0.064 と低く、この点からも betel-chewing と口腔がんの関連は低いといえる。