

R Code And Tasks Chapter 5 (MAS 6003)

Witold Wolski

December 28, 2016

6.1 Types of 2 way tables - response / control variables

6.1.1 Case(a): Skin cancer (melanoma) data - 2 response variables

Cross sectional study of malignant melanoma. Both tumour type and site are **response variables** because none of the row or column totals were fixed in advance of the data collection.

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.3.2
```

```
rm(list=ls())
load("data/MAS367-GLMs.RData", envir = e <- new.env())

Mela <- e$Mela
head(Mela)
```

```
##   number tumour.type      site
## 1     22          A Head.and.Neck
## 2     16          B Head.and.Neck
## 3     19          C Head.and.Neck
## 4     11          D Head.and.Neck
## 5      2          A      Trunk
## 6     54          B      Trunk
```

```
dcast(Mela,tumour.type~ site, value.var="number")
```

```
##   tumour.type Head.and.Neck Trunk Extremities
## 1          A           22     2           10
## 2          B           16    54          115
## 3          C           19    33           73
## 4          D           11    17           28
```

6.1.2 Case(b) : Flu vaccine data - 1 response and 1 control variable

Patients were randomly assigned to the two groups (Placebo, Vaccine), and the response (levels of an antibody found in the blood six weeks after vaccination) was determined. Antibody level is the **response** and vaccine group is a **controlled variable** (with totals fixed by experimental design).

```
vaccine <- e$vaccine
head(vaccine)
```

```
##   count response treatment
## 1    25    small  placebo
## 2     6    small  vaccine
## 3     8 moderate placebo
## 4    18 moderate vaccine
## 5     5    large placebo
## 6    11    large  vaccine
```

```
dcast(vaccine, treatment ~ response, value.var="count")
```

```
##   treatment small moderate large
## 1   placebo    25         8     5
## 2   vaccine     6        18    11
```

6.2.1 Association, Independence and Homogeneity

Independence :

Case(a)

Probabilities of interest $\pi_{ij} = P(A = i, B = j)$

$$P(A = i, B = j) = P(A = i) \times P(B = j)$$

Case(b)

Probabilities of interest are conditional probabilities $\pi_{ij} = P(B = j | A = i)$.

The interest is in whether the probability distribution of the response (antibody level) is the same in each level of the controlled variable (drug group). If it doesn't depend on i then we can write

$$(\pi_{ij} = \pi_{.j})$$

.

This is known as **homogeneity**.

6.3 Distribution for two-way tables

6.3.1 Case(a): two response variables

6.3.2 Case(b): one response variable

6.3.3 Case(c): independent poisson (no fixed margins).

6.3.4 Expected values

6.4 GLMs and two-way contingency tables

6.4.1 Natural hypothesis are log-linear models

6.4.2 Poisson log-linear modelling for two-way tables

6.4.3 Maximum likelihood estimation for π_{ij} in case (a) and Task 19 and Task 20.

Task 19 Verify the maximum likelihood estimate for π_{ij} for the A + B model for case(a). Task 20 Verify the maximum likelihood estimate for π_{ij} for the A + B model for case(b).

6.5 Interaction plots and examination of residues.

`interaction.plot` honorable mentioned.