

Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

R Demo for Chapter 15, part 1: Contingency Tables

Jinko Graham

Contingency Tables

- ▶ Example: Data of Mungan et al. 2000 click on 21,737 bladder cancer patients
 - ▶ Two categorical variables: **gender**, which has 2 levels, and **cancer stage**, which has 4 levels.

```
uu <- url("http://people.stat.sfu.ca/~jgraham/Teaching/S305_17/Data/mung.csv")
Mungan <- read.csv(uu)
head(Mungan,n=8)
```

```
##   Gender Cancer.Stage
## 1   Male           I
## 2   Male           I
## 3   Male           I
## 4   Male           I
## 5   Male           I
## 6   Male           I
## 7   Male           I
## 8   Male           I
```

Tabulating Data in R

- ▶ Two useful functions in R for cross tabulating:
 - ▶ `table()` - easiest to use and
 - ▶ `xtabs()` - more flexible, but flexibility only needed when we have more than two variables – see Chapter 16

```
table(Mungan)
```

```
##           Cancer.Stage
## Gender      I      II   III   IV
## Female  3926   402   356   852
## Male   12418   995   883  1905
```

Row and Column Margins

- ▶ In R, the **row margin** is the tabulation of the row variable and the **column margin** is the tabulation of the column variable.
 - ▶ The row margin is indexed by the number 1 and the column margin by the number 2.

```
mtab <- table(Mungan)
margin.table(mtab,1) # R's row margin
```

```
## Gender
## Female    Male
##    5536   16201
```

```
margin.table(mtab,2) # R's column margin
```

```
## Cancer.Stage
##      I      II     III     IV
## 16344  1397  1239   2757
```

Cancer stage distribution given gender

```
mtab <- table(Mungan)
mtab
```

```
##           Cancer.Stage
## Gender           I      II      III      IV
## Female    3926    402    356    852
## Male     12418    995    883   1905
```

- For each gender category, we can divide the counts in each row by the row total to get proportions.

```
prop.table(mtab,margin=1) # margin=1 refers to dividing by row totals
```

```
##           Cancer.Stage
## Gender           I           II           III           IV
## Female 0.70917630 0.07261561 0.06430636 0.15390173
## Male   0.76649590 0.06141596 0.05450281 0.11758533
```

Gender distribution given cancer stage

- Likewise, for each cancer stage category we can divide the counts in each column by the column total to get proportions.

```
prop.table(mtab,margin=2)
```

```
##           Cancer.Stage
## Gender           I           II           III           IV
##   Female 0.2402105 0.2877595 0.2873285 0.3090316
##   Male   0.7597895 0.7122405 0.7126715 0.6909684
```