# STA312f22 Assignment 5



① (a) $G^2 = -2 \log \frac{\ell(p_1, p_2)}{\ell(\hat{\pi}_1, \hat{\pi}_2)}$

If $\pi_1 = \pi_2 = \pi$, $\frac{\partial}{\partial \pi} \log \ell$

$= \frac{\partial}{\partial \pi} \log\left[ \prod_{i=1}^{n_1} \pi^{x_i} (1-\pi)^{1-x_i} \prod_{j=1}^{n_2} \pi^{y_j} (1-\pi)^{1-y_j} \right]$

$= \frac{\partial}{\partial \pi} \log\left[ \pi^{\Sigma x + \Sigma y} (1-\pi)^{n_1+n_2-\Sigma x - \Sigma y} \right]$

$= \frac{\partial}{\partial \pi} \left( (\Sigma x + \Sigma y) \log \pi + (n_1+n_2-\Sigma x - \Sigma y) \log(1-\pi) \right)$

$= \frac{\Sigma x + \Sigma y}{\pi} - \frac{n_1+n_2-\Sigma x - \Sigma y}{1-\pi} \stackrel{set}{=} 0$

$\iff \pi(n_1+n_2-\Sigma x - \Sigma y) = \Sigma x + \Sigma y - \pi(\Sigma x + \Sigma y)$

$\iff \pi(n_1+n_2 - \Sigma x - \Sigma y + \Sigma x + \Sigma y) = \Sigma x + \Sigma y$

$\implies \hat{\pi} = \frac{\Sigma x + \Sigma y}{n_1+n_2} = \hat{\pi}$

(1a cont)

$$G^2 = -2 \log \frac{\hat{\pi}^{\Sigma x}(1-\hat{\pi})^{n_1-\Sigma x}\,\hat{\pi}^{\Sigma y}(1-\hat{\pi})^{n_2-\Sigma y}}{p_1^{\Sigma x}(1-p_1)^{n_1-\Sigma x}\,p_2^{\Sigma y}(1-p_2)^{n_2-\Sigma y}}$$

$$= 2\log\left[\left(\frac{p_1}{\hat{\pi}}\right)^{\Sigma x}\left(\frac{1-p_1}{1-\hat{\pi}}\right)^{n_1-\Sigma x}\left(\frac{p_2}{\hat{\pi}}\right)^{\Sigma y}\left(\frac{1-p_2}{1-\hat{\pi}}\right)^{n_2-\Sigma y}\right]$$

$$= 2\left(\Sigma x \log\frac{p_1}{\hat{\pi}} + (n_1-\Sigma x)\log\left(\frac{1-p_1}{1-\hat{\pi}}\right) + \Sigma y \log\frac{p_2}{\hat{\pi}} + (n_2-\Sigma y)\log\left(\frac{1-p_2}{1-\hat{\pi}}\right)\right)$$

(b) $p_1 = 0.221,\ p_2 = 0.122,\ n_1 = 140,\ n_2 = 139$

$\Sigma x = p_1 n_1 = 0.221 * 140 = 30.94 \rightarrow 31$

$\Sigma y = p_2 n_2 = 0.112 * 139 = 16.958 \rightarrow 17$

$n_1 - \Sigma x = 140 - 31 = 109,\ n_2 - \Sigma y = 139 - 17 = 122$

$\hat{\pi} = \frac{31+17}{140+139} = \frac{48}{279} = 0.172$

$$G^2 = 2\left(31\log\frac{0.221}{0.172} + 109\log\frac{0.779}{0.828} + 17\log\frac{0.122}{0.172} + 122\,\frac{0.878}{0.828}\right)$$

$$= 2(7.77 - 6.65 - 5.84 + 7.15) = 2*2.43$$

$= 4.86,\ df = 1,\ p = 0.027$

(c)

| | Sick | Not | |
|---|---|---|---|
| Placebo | 31 | 109 | 140 |
| Drug | 17 | 122 | 139 |

$G^2 = 4.87$ in the lecture slides (Contingency Tables with R)

(2) (a) Two; a ≠ b

(b) $\ell_0 = (ab)^{n_{11}} (a(1-b))^{n_{12}} ((1-a)b)^{n_{21}} ((1-a)(1-b))^{n_{22}}$

$= a^{n_{11}+n_{12}} b^{n_{11}+n_{21}} (1-a)^{n_{21}+n_{22}} (1-b)^{n_{12}+n_{22}}$

(c) $\frac{\partial}{\partial a} \log \ell_0$

$= \frac{\partial}{\partial a} \Big( (n_{11}+n_{12}) \log a + (n_{11}+n_{21}) \log b + (n_{21}+n_{22}) \log (1-a) + (n_{12}+n_{22}) \log (1-b) \Big)$

$= \frac{n_{11}+n_{12}}{a} - \frac{n_{21}+n_{22}}{1-a} \overset{set}{=} 0$

| | |
|---|---|
| $n_{11}$ | $n_{12}$ |
| $n_{21}$ | $n_{22}$ |

$\Rightarrow n_{11}+n_{12} - a(n_{11}+n_{12}) = a(n_{21}+n_{22})$

$\Rightarrow n_{11}+n_{12} = a(n_{11}+n_{12}+n_{21}+n_{22}) = an$

$\Rightarrow \hat{a} = \frac{n_{11}+n_{12}}{n} = \frac{n_{1+}}{n}$

$\frac{\partial}{\partial b} \log \ell_0 = \frac{n_{11}+n_{21}}{b} - \frac{n_{12}+n_{22}}{1-b} \overset{set}{=} 0$

$\Rightarrow n_{11}+n_{21} - b(n_{11}+n_{21}) = b(n_{12}+n_{22})$ SAME AS BEFORE

$\Rightarrow \hat{b} = \frac{n_{11}+n_{21}}{n} = \frac{n_{+1}}{n}$

(2)

(d)

| | | | |
|---|---|---|---|
| | $\frac{n_{1+}}{n}\frac{n_{+1}}{n}$ | $\frac{n_{1+}}{n}\frac{n_{+2}}{n}$ | $n_{1+}$ |
| | $\frac{n_{+1}}{n}\frac{n_{2+}}{n}$ | $\frac{n_{2+}}{n}\frac{n_{+2}}{n}$ | $n_{2+}$ |
| | $n_{+1}$ | $n_{+2}$ | |

multiply by $n$ to get expected values

(e) $\hat{\mu}_{ij} = \frac{n_{i+} n_{+j}}{n}$

(3) (a) $d = \frac{\pi}{1-\pi} \Rightarrow d - \pi d = \pi \Rightarrow d = \pi(1+d)$

$$\Rightarrow \pi = \frac{d}{1+d}$$

(b) Use the quotient rule I guess

$$\frac{d}{d\pi}\left(\frac{\pi}{1-\pi}\right) = \frac{u'v - v'u}{v^2} = \frac{1-\pi-(-1)\pi}{(1-\pi)^2}$$

$$= \frac{1-\pi+\pi}{(1-\pi)^2} = \frac{1}{(1-\pi)^2} > 0$$

increasing

(4) (a) $\dfrac{\pi_{11}}{\pi_{11}+\pi_{12}} \Big/ \dfrac{\pi_{12}}{\pi_{11}+\pi_{12}} = \dfrac{\pi_{11}}{\pi_{12}}$

(b) $\dfrac{\pi_{21}}{\pi_{22}}$

(c) $\dfrac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \dfrac{\pi_{11}\,\pi_{22}}{\pi_{12}\,\pi_{21}} = \theta$

(d) ~~First assume $P(Y=1|X=1)=P(Y=1|X=2)$ & show $\theta=1$~~

$$\frac{\pi_{11}}{\pi_{11}+\pi_{12}} = \frac{\pi_{21}}{\pi_{21}+\pi_{22}} \iff$$

$$\pi_{11}\pi_{21} + \pi_{11}\pi_{22} = \pi_{11}\pi_{21} + \pi_{12}\pi_{21}$$

$$\iff \frac{\pi_{11}\,\pi_{22}}{\pi_{12}\,\pi_{21}}$$ And arrows go in both directions

(e) $P(X=1|Y=1) = P(X=1|Y=2) \iff \dfrac{\pi_{11}}{\pi_{11}+\pi_{21}} = \dfrac{\pi_{12}}{\pi_{12}+\pi_{22}}$

$$\iff \pi_{11}\pi_{12} + \pi_{11}\pi_{22} = \pi_{11}\pi_{12} + \pi_{12}\pi_{21}$$

$$\iff \frac{\pi_{11}\pi_{22}}{\pi_{12}\,\pi_{21}} = 1 \iff \theta = 1$$

(f) $\begin{vmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{vmatrix} = \pi_{11}\pi_{22} - \pi_{12}\pi_{21} = 0 \iff \pi_{11}\pi_{22} = \pi_{12}\pi_{21}$

$\iff \theta = 1$

(5)

| | $Y=1$ | $Y=2$ |
|---|---|---|
| $X=1$ | $\pi_{11}$ | $1-\pi_{11}$ |
| $X=2$ | $\pi_{21}$ | $1-\pi_{21}$ |

$$\text{odds ratio} = \frac{\pi_{11}/(1-\pi_{11})}{\pi_{21}/(1-\pi_{21})} = \frac{\pi_{11}(1-\pi_{21})}{\pi_{21}(1-\pi_{11})} = \theta$$

(6)

| | $Y=1$ | $Y=2$ |
|---|---|---|
| $X=1$ | $\pi_{11}$ | $\pi_{12}$ |
| $X=2$ | $1-\pi_{11}$ | $1-\pi_{12}$ |

$$\text{odds ratio} = \frac{\pi_{11}/(1-\pi_{11})}{\pi_{12}/(1-\pi_{12})} = \frac{\pi_{11}(1-\pi_{12})}{\pi_{12}(1-\pi_{11})} = \theta$$

7 (a) $\binom{n}{a}\binom{n}{b}$

(b) $\binom{n}{x \quad a-x \quad b-x \quad n-a-b+x}$

(c) $$P(n_{11}=x) = \frac{\dfrac{n!}{x!\,(a-x)!\,(b-x)!\,(n-a-(b-x))!}}{\dfrac{n!}{a!\,(n-a)!}\binom{n}{b}}$$

$$= \frac{\dfrac{a!}{x!\,(a-x)!}\cdot\dfrac{(n-a)!}{(b-x)!\,((n-a)-(b-x))!}}{\binom{n}{b}}$$

$$= \frac{\binom{a}{x}\binom{n-a}{b-x}}{\binom{n}{b}}$$

(7d)
$$\theta = \frac{x(n-a-b+x)}{(a-x)(b-x)}$$

$$\log\theta = \log x + \log(n-a-b+x) - \log(a-x) - \log(b-x)$$

$$\frac{d}{dx}\log\theta = \frac{1}{x} + \frac{1}{n-a-b+x} - \frac{1}{a-x}(-1) - \frac{1}{b-x}(-1)$$

$$= \frac{1}{x} + \frac{1}{n-a-b+x} + \frac{1}{a-x} + \frac{1}{b-x} > 0$$

(e) $x \geq 0$, $a-x \geq 0 \iff x \leq a$

$b-x \geq 0 \iff x \leq b$

$n-a-b+x \geq 0 \iff x \geq a+b-n$, so

$$\text{Max}(0, a+b-n) \leq x \leq \text{Min}(a, b)$$

```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.78 (8075) x86_64-apple-darwin17.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

> # R work for Titanic Question 8.
>
> # help(Titanic)
> # No Name    Levels
> # 1  Class       1st, 2nd, 3rd, Crew
> # 2  Sex         Male, Female
> # 3  Age         Child, Adult
> # 4  Survived     No, Yes
>
> # No crew, adults only.
>
> # Odds ratio function
> oddrat = function(M) M[1,1]*M[2,2] / (M[1,2]*M[2,1])
>
> ship = as.data.frame(Titanic) ; ship
   Class    Sex   Age Survived Freq
1    1st   Male Child       No    0
2    2nd   Male Child       No    0
3    3rd   Male Child       No   35
4   Crew   Male Child       No    0
5    1st Female Child       No    0
6    2nd Female Child       No    0
7    3rd Female Child       No   17
8   Crew Female Child       No    0
9    1st   Male Adult       No  118
10   2nd   Male Adult       No  154
11   3rd   Male Adult       No  387
12  Crew   Male Adult       No  670
13   1st Female Adult       No    4
14   2nd Female Adult       No   13
15   3rd Female Adult       No   89
16  Crew Female Adult       No    3
17   1st   Male Child      Yes    5
18   2nd   Male Child      Yes   11
```

```
19   3rd   Male Child      Yes   13
20  Crew   Male Child      Yes    0
21   1st Female Child      Yes    1
22   2nd Female Child      Yes   13
23   3rd Female Child      Yes   14
24  Crew Female Child      Yes    0
25   1st   Male Adult      Yes   57
26   2nd   Male Adult      Yes   14
27   3rd   Male Adult      Yes   75
28  Crew   Male Adult      Yes  192
29   1st Female Adult      Yes  140
30   2nd Female Adult      Yes   80
31   3rd Female Adult      Yes   76
32  Crew Female Adult      Yes   20
> iceberg1 = subset(ship, Age == 'Adult') # Select adults
> iceberg1 = subset(iceberg1, Class != 'Crew')
> iceberg1
   Class    Sex   Age Survived Freq
9    1st   Male Adult       No  118
10   2nd   Male Adult       No  154
11   3rd   Male Adult       No  387
13   1st Female Adult       No    4
14   2nd Female Adult       No   13
15   3rd Female Adult       No   89
25   1st   Male Adult      Yes   57
26   2nd   Male Adult      Yes   14
27   3rd   Male Adult      Yes   75
29   1st Female Adult      Yes  140
30   2nd Female Adult      Yes   80
31   3rd Female Adult      Yes   76
>
> # 8a) Cross-sectional
>
> # 8b)
> sexBYsurv = xtabs(Freq ~ Sex + Survived, data = iceberg1)
> sexBYsurv
        Survived
Sex       No Yes
  Male   659 146
  Female 106 296
> round(prop.table(sexBYsurv,margin=1),3)
        Survived
Sex         No   Yes
  Male   0.819 0.181
  Female 0.264 0.736
> oddrat(sexBYsurv)
[1] 12.60429
> # The odds of death were 12.6 times as great for men.
> chisq.test(sexBYsurv,correct=FALSE)

	Pearson's Chi-squared test

data:  sexBYsurv
X-squared = 355.76, df = 1, p-value < 2.2e-16

> # Women wee more likely to survive than men.
>
>
```

```
>
> # 8c)
> classBYsexBYsurv = xtabs(Freq ~ Sex + Survived + Class, data = iceberg1)
> classBYsexBYsurv
, , Class = 1st

        Survived
Sex       No Yes
  Male   118  57
  Female   4 140

, , Class = 2nd

        Survived
Sex       No Yes
  Male   154  14
  Female  13  80

, , Class = 3rd

        Survived
Sex       No Yes
  Male   387  75
  Female  89  76

, , Class = Crew

        Survived
Sex       No Yes
  Male     0   0
  Female   0   0

> # Get rid of empty crew.
> classBYsexBYsurv = classBYsexBYsurv[,,(1:3)]
> classBYsexBYsurv
, , Class = 1st

        Survived
Sex       No Yes
  Male   118  57
  Female   4 140

, , Class = 2nd

        Survived
Sex       No Yes
  Male   154  14
  Female  13  80

, , Class = 3rd

        Survived
Sex       No Yes
  Male   387  75
  Female  89  76

> # Show them how to do it later by applying factor after selection.
>
```

```
> round(prop.table(classBYsexBYsurv,margin=c(1,3)),3)
, , Class = 1st

        Survived
Sex          No   Yes
  Male   0.674 0.326
  Female 0.028 0.972

, , Class = 2nd

        Survived
Sex          No   Yes
  Male   0.917 0.083
  Female 0.140 0.860

, , Class = 3rd

        Survived
Sex          No   Yes
  Male   0.838 0.162
  Female 0.539 0.461

> # Chi-squared separately for each class.
> for(j in 1:3) print(chisq.test(classBYsexBYsurv[,,j]))

	Pearson's Chi-squared test with Yates' continuity correction

data:  classBYsexBYsurv[, , j]
X-squared = 137.08, df = 1, p-value < 2.2e-16


	Pearson's Chi-squared test with Yates' continuity correction

data:  classBYsexBYsurv[, , j]
X-squared = 153.43, df = 1, p-value < 2.2e-16


	Pearson's Chi-squared test with Yates' continuity correction

data:  classBYsexBYsurv[, , j]
X-squared = 57.539, df = 1, p-value = 3.313e-14

>
> for(j in 1:3) cat('Odds ratio for class ',j,' = ',
+ oddrat(classBYsexBYsurv[,,j]),'\n')
Odds ratio for class  1  =  72.45614
Odds ratio for class  2  =  67.69231
Odds ratio for class  3  =  4.406292
>
> # Try children.
> # Class 1-3, All sex, 1=children, All survival outcomes
> Titanic[(1:3),,1,]
, , Survived = No

     Sex
Class Male Female
  1st    0      0
  2nd    0      0
```

```
  3rd   35     17

, , Survived = Yes

     Sex
Class Male Female
  1st    5      1
  2nd   11     13
  3rd   13     14

> iceberg2 = subset(ship, Age == 'Child') # Select children
> iceberg2
   Class    Sex   Age Survived Freq
1    1st   Male Child       No    0
2    2nd   Male Child       No    0
3    3rd   Male Child       No   35
4   Crew   Male Child       No    0
5    1st Female Child       No    0
6    2nd Female Child       No    0
7    3rd Female Child       No   17
8   Crew Female Child       No    0
17   1st   Male Child      Yes    5
18   2nd   Male Child      Yes   11
19   3rd   Male Child      Yes   13
20  Crew   Male Child      Yes    0
21   1st Female Child      Yes    1
22   2nd Female Child      Yes   13
23   3rd Female Child      Yes   14
24  Crew Female Child      Yes    0
>
> kidstable = xtabs(Freq ~ Sex + Survived + Class, data = iceberg2)
> kidstable
, , Class = 1st

        Survived
Sex       No Yes
  Male     0   5
  Female   0   1

, , Class = 2nd

        Survived
Sex       No Yes
  Male     0  11
  Female   0  13

, , Class = 3rd

        Survived
Sex       No Yes
  Male    35  13
  Female  17  14

, , Class = Crew

        Survived
Sex       No Yes
  Male     0   0
```

```
  Female  0   0

> poorkids = kidstable[,,3]; poorkids
        Survived
Sex      No Yes
  Male   35  13
  Female 17  14
> round(prop.table(poorkids,margin=1),3)
        Survived
Sex         No   Yes
  Male   0.729 0.271
  Female 0.548 0.452
> oddrat(poorkids)
[1] 2.217195
> chisq.test(poorkids,correct=FALSE)

    Pearson's Chi-squared test

data:  poorkids
X-squared = 2.7363, df = 1, p-value = 0.09809

>
> # In first and second class, all the children survived. In third class, there is no convincing
evidence of a difference in survival rate between boys and girls.
>
>
```