Qualitative Predictors: Passive Smoking Revisit

Spouse Smoked	Japan Case	Japan Control	UK Case	UK Control	US Case	US Control
Yes	73	188	19	38	137	363
No	21	82	5	16	71	249

Model: $\text{logit}(\pi) = \alpha + \beta x + \beta_{UK} C_{UK} + \beta_{US} C_{US}$

$\pi = \text{P}(\text{Case (lung cancer)})$

$$x = \begin{cases} 1 & \text{if passive smoking} \\ 0 & \text{if no passive smoking} \end{cases}$$

$$C_{UK} = \begin{cases} 1 & \text{if Country = UK} \\ 0 & \text{if Country = JP or US} \end{cases}$$

$$C_{US} = \begin{cases} 1 & \text{if Country = US} \\ 0 & \text{if Country = JP or UK} \end{cases}$$

Country	Passive Smoking	logit(π)
JP	N	α
	Y	$\alpha + \beta$
UK	N	$\alpha + \quad + \beta_{UK}$
	Y	$\alpha + \beta + \beta_{UK}$
US	N	$\alpha + \quad + \beta_{US}$
	Y	$\alpha + \beta + \beta_{US}$

Homogeneous Association

The model

$$\text{logit}(\pi) = \alpha + \beta x + \beta_{UK} C_{UK} + \beta_{US} C_{US}$$

has **no interaction term**, which means the same conditional odds ratio

$$\frac{\text{odds for passive smokers}}{\text{odds for non-passive smokers}} = \frac{e^{\alpha+\beta+\beta_{UK}C_{UK}+\beta_{US}C_{US}}}{e^{\alpha+\beta_{UK}C_{UK}+\beta_{US}C_{US}}} = e^{\beta}$$

for both levels of initial size of stone. That is **homogeneous association** — same conditional odds ratio at each level of other variable.

Likewise, the conditional odds ratio for "Country" is also constant regardless of smoking status.

$$\frac{\text{odds for UK}}{\text{odds for JP}} = \frac{e^{\alpha+\beta x+\beta_{UK}}}{e^{\alpha+\beta x}} = e^{\beta_{UK}}$$


```
> Case = c(73, 21, 19, 5, 137, 71)
> Control = c(188, 82, 38, 16, 363, 249)
> SpouseSmoking = rep(c("Yes","No"), 3)
> Country = c("JP","JP","UK","UK","US","US")
> PassSmok = data.frame(SpouseSmoking, Country, Case, Control)
> PassSmok
  SpouseSmoking Country Case Control
1           Yes      JP   73     188
2            No      JP   21      82
3           Yes      UK   19      38
4            No      UK    5      16
5           Yes      US  137     363
6            No      US   71     249
```



```
> fit1 = glm(cbind(Case, Control) ~ Country + SpouseSmoking,
             family = binomial, data=PassSmok)
> summary(fit1)
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.293807   0.159199  -8.127  4.4e-16 ***
CountryUK        0.240844   0.273559   0.880   0.3786
CountryUS        0.009867   0.145148   0.068   0.9458
SpouseSmokingYes 0.325530   0.139590   2.332   0.0197 *
```

After accounting for country effect, odds of getting lung cancer for passive smokers are estimated to be $e^{\widehat{\beta}} = e^{0.3255} \approx 1.38$ times the odds for non-passive smokers.

95% Wald CI for e^{β}:

$$e^{\widehat{\beta}\pm 1.96\times \text{SE}} = e^{0.3255\pm 1.96\times 0.1396} = (e^{0.052}, e^{0.599}) \approx (1.05, 1.82)$$

Significant adverse effect of passive smoking after accounting for country effect.

Tests of Conditional Independence

In the model

$$\text{logit}(\pi) = \alpha + \beta x + \beta_{UK} C_{UK} + \beta_{US} C_{US},$$

$\beta = 0$ means conditional odds ratio $e^{\beta} = e^{0} = 1$, i.e., lung cancer and passive smoking are **conditionally independent** given country.

Tests of conditional independence:

- CMH test
  - In fact, CMH test is the score test of $\beta = 0$ in the logistic model
- Wald test of $\beta = 0$ in the logistic model
- LR test of $\beta = 0$ in the logistic model

Tests of Conditional Independence (Cont'd)

Wald test of conditional independence gives P-value $= 0.0197$

```
> summary(fit1)
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.293807   0.159199  -8.127  4.4e-16 ***
CountryUK         0.240844   0.273559   0.880   0.3786
CountryUS         0.009867   0.145148   0.068   0.9458
SpouseSmokingYes  0.325530   0.139590   2.332   0.0197 *
```

LR test of conditional independence gives P-value $= 0.01842$:

```
> drop1(fit1, test="Chisq")
Single term deletions

Model:
cbind(Case, Control) ~ Country + SpouseSmoking
              Df Deviance    AIC    LRT Pr(>Chi)
<none>            0.2396 38.595
Country        2   1.0647 35.420 0.8251  0.66195
SpouseSmoking  1   5.7952 42.150 5.5556  0.01842 *
```

CMH test gives the P-value 0.01957 (See Slide `C02D.pdf`).

Comparison of the Three Tests of Conditional Independence

- The three tests usually agree when the sample sizes in each partial table are big enough
- Wald and LR tests require the sample size in each partial table to be large enough
- CMH test can work when the counts in the partial tables are small as long as the overall count is large enough
- In H_a, Wald and LR tests assume homogeneous association, but CMH test does not assume equality of odds ratios
- To sum up, for testing conditional independence in $2 \times 2 \times K$ tables, CMH test is preferred over Wald or LR tests.

Estimation of Common Odds Ratio

- MH estimate of the common odds ratio (See Slide `C02D.pdf`).
- In the logistic regression model:

  $$\text{logit}(\pi) = \alpha + \beta x + \beta_{UK} C_{UK} + \beta_{US} C_{US},$$

  e^{β} is the common odds ratio, and $e^{\widehat{\beta}}$ is the maximum likelihood estimate (MLE) for the common odds ratio. One can construct the Wald or LR confidence interval for e^{β}
- MH estimate is preferred over MLE of the common odds ratio.

Test of Homogeneous Association

If we include the interaction term,

Model 2: $\text{logit}(\pi) = \alpha+\beta x+\beta_{UK} C_{UK}+\beta_{US} C_{US}+\gamma_{UK} x C_{UK}+\gamma_{US} x C_{US}$,

the conditional odds ratio

$$\frac{\text{odds for Passive Smokers}}{\text{odds for Non-Passive Smokers}} = \frac{e^{\alpha+\beta+\beta_{UK} C_{UK}+\beta_{US} C_{US}+\gamma_{UK} C_{UK}+\gamma_{US} C_{US}}}{e^{\alpha+\beta_{UK} C_{UK}+\beta_{US} C_{US}}} = e^{\beta+\gamma_{UK} C_{UK}+\gamma_{US} C_{US}}$$

changes with Country, if γ_{UK} or $\gamma_{US} \neq 0$.

H_0: $\gamma_{UK} = \gamma_{US} = 0$ means homogeneous association.

```
> fit2 = glm(cbind(Case, Control) ~ Country + SpouseSmoking + Country:SpouseSmoking,
             family = binomial, data=PassSmok)
> anova(fit1, fit2, test="Chisq")
Analysis of Deviance Table

Model 1: cbind(Case, Control) ~ Country + SpouseSmoking
Model 2: cbind(Case, Control) ~ Country + SpouseSmoking + Country:SpouseSmoking
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         2    0.23958
2         0    0.00000  2  0.23958   0.8871
```