More multinomial regression

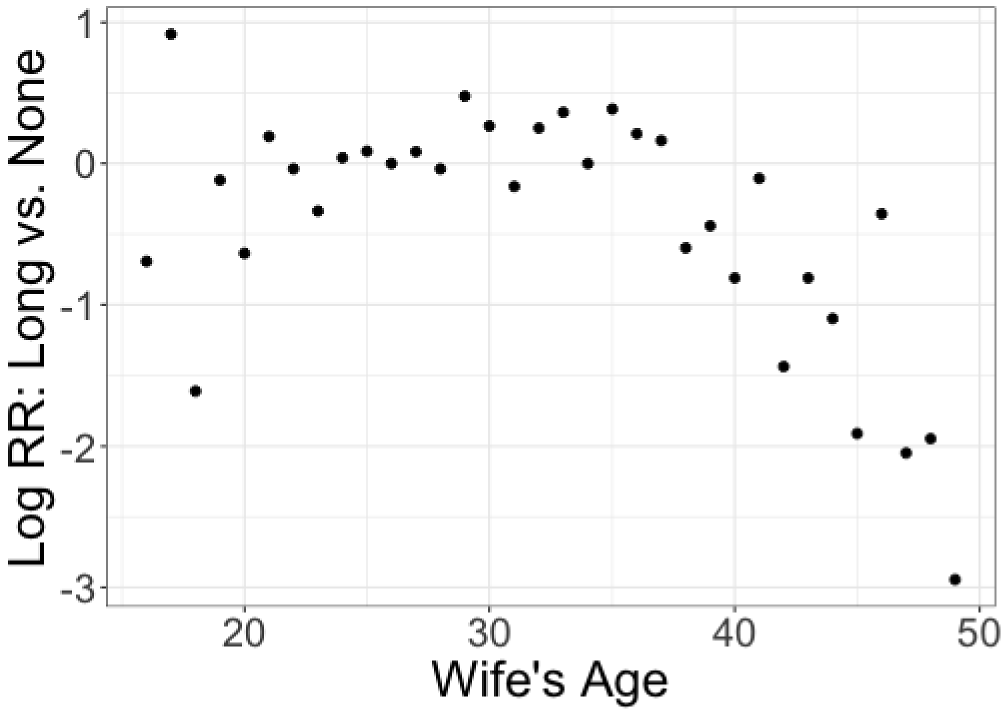
Motivation

Question: What is the relationship between age and contraceptive use for women in Indonesia?

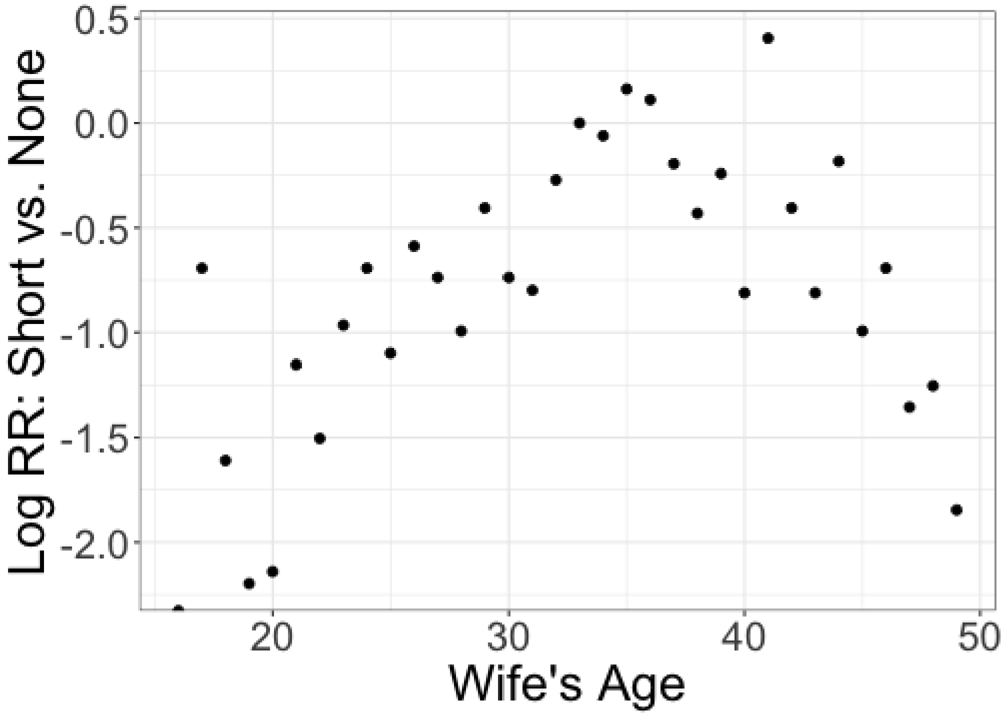
Data: 1473 Indonesian couples, with variables

- Y_i = contraceptive method used (1 = no use, 2 = long-term, 3 = short-term)
- X_i = Wife's age (numeric)

EDA: log relative risk



EDA: log relative risk



Multinomial regression model

Step 1: Choose a reasonable distribution for Y

$$Y_i \sim Categorical(\pi_{i(None)}, \pi_{i(Short)}, \pi_{i(Long)})$$

Step 2: Choose a model for any parameters

$$\log\left(\frac{\pi_{i(Long)}}{\pi_{i(None)}}\right) = \beta_{0(Long)} + \beta_{1(Long)} Age_i + \beta_{2(Long)} Age_i^2$$

$$\log\left(\frac{\pi_{i(Short)}}{\pi_{i(None)}}\right) = \beta_{0(Short)} + \beta_{1(Short)} Age_i + \beta_{2(Short)} Age_i^2$$

Fitting the model in R

- To fit a multinomial regression, we use the `nnet` package

```
library(nnet)
```

- Syntax is very similar to other regression techniques

```
m1 <- multinom(Choice ~ WifeAge + I(WifeAge^2),
               data = cmc_data)
```

Fitting the model in R

When you fit the model, you get the following output:

```
m1 <- multinom(Choice ~ WifeAge + I(WifeAge^2),
               data = cmc_data)
```

```
## # weights:  12 (6 variable)
## initial  value 1618.255901
## iter  10 value 1507.910585
## iter  10 value 1507.910578
## iter  10 value 1507.910578
## final  value 1507.910578
## converged
```

- This reflects the process used to estimate the model parameters -- we won't get into that here

Fitting the model in R

```
summary(m1)
```

```
## Call:
## multinom(formula = Choice ~ WifeAge + I(WifeAge^2), data
##
## Coefficients:
##       (Intercept)   WifeAge I(WifeAge^2)
## Short   -8.234242 0.4562421 -0.006462919
## Long    -5.083101 0.3656354 -0.006279489
##
## Std. Errors:
##        (Intercept)    WifeAge I(WifeAge^2)
## Short 0.0005964133 0.01009311 0.0002632050
## Long  0.0005593739 0.00878999 0.0002471778
##
## Residual Deviance: 3015.821
## AIC: 3027.821
```

Fitting the model in R

```
...
## Coefficients:
##       (Intercept)   WifeAge I(WifeAge^2)
## Short   -8.234242 0.4562421 -0.006462919
## Long    -5.083101 0.3656354 -0.006279489
...
```

> What is the fitted model for short term vs. no contraceptive use?

Class activity, Part I

https://sta214-f22.github.io/class_activities/ca_lecture_15.html

Class activity

> What is the predicted relative risk of short term use vs. no use use for a woman aged 25?

Class activity

What is the predicted *probability* of each contraceptive choice for a woman aged 25?

Generalizing

Let $Y_i \sim Categorical(\pi_{i1}, \ldots, \pi_{iJ})$ be a categorical variable with J levels, and let j^* be the reference level. For each $\pi_{ij}, j \neq j^*$, we model

$$\log\left(\frac{\pi_{ij}}{\pi_{ij^*}}\right) = \beta_{0(j)} + \beta_{1(j)} X_i$$

Then $\pi_{ij} = \pi_{ij^*} \exp\{\beta_{0(j)} + \beta_{1(j)} X_i\}$, so

$$\pi_{ij^*} = \frac{1}{1 + \sum_{j \neq j^*} \exp\{\beta_{0(j)} + \beta_{1(j)} X_i\}}$$

$$\pi_{ij} = \frac{\exp\{\beta_{0(j)} + \beta_{1(j)} X_i\}}{1 + \sum_{j \neq j^*} \exp\{\beta_{0(j)} + \beta_{1(j)} X_i\}}$$

Predicted probabilities in R

We can obtain the predicted probabilities for each individual in the data:

```
probspred <- fitted(m1)
```

Here are the predicted probabilities for the second individual:

```
probspred[2,]
```

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

Making predictions

```
probspred[2,]
```

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

This is a 24 year old woman who does not use contraceptives.

> How well did we do at estimating their chances of using contraception?

Making predictions

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

This is a 24 year old woman who does not use contraceptives.

> What would our predicted probabilities be if we were just guessing?

Making predictions

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

This is a 24 year old woman who does not use contraceptives.

- If we don't have any data, our estimated probability would be 1/3 for each level
- If we have data but we don't use age, our estimated probability for each level is just the proportion of observations in that group:

```
table(cmc_data$Choice)/nrow(cmc_data)
```

```
##
##      None     Short      Long
## 0.4270197 0.2260692 0.3469111
```

Making predictions

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

This is a 24 year old woman who does not use contraceptives.

> For this individual, are we doing better than random guessing?

Making predictions

Given predicted probabilities $\widehat{\pi}_{ij}$ for individual i, how could I predict the response $\widehat{Y}_i$?

Making predictions

```
probspred <- fitted(m1)
probspred[2,]
```

```
##      None     Short      Long
## 0.5834588 0.2645993 0.1519420
```

```
preds <- predict(m1)
preds[2]
```

```
## [1] None
## Levels: None Short Long
```

Making predictions

How can I assess all of my predictions at once?

Making predictions

> How can I assess all of my predictions at once?

With a confusion matrix!

```r
table("Prediction" = preds,
      "Actual" = cmc_data$Choice)
```

```
##             Actual
## Prediction None Short Long
##      None   342   166  189
##      Short    0     0    0
##      Long   287   167  322
```

Making predictions

```
##             Actual
## Prediction None Short Long
##       None  342   166  189
##      Short    0     0    0
##       Long  287   167  322
```

> Does it look like we're doing a good job at predicting contraception use?

Class activity, Part II

https://sta214-f22.github.io/class_activities/ca_lecture_15.html

Class activity

```
##              Actual
## Prediction None Short Long
##       None   342   166  189
##       Short    0     0    0
##       Long   287   167  322
```

What fraction of our predictions are correct?

Class activity

> What would our confusion matrix look like if our predictions randomly assigned each person to one of the three categories, with a 1/3 chance for each category?

Class activity

> What would our confusion matrix look like if our predictions randomly assigned each person to one of the three categories, with a 1/3 chance for each category?

Something like

		Actual		
		None	Short	Long
Predicted	None	210	111	170
	Short	210	111	170
	Long	209	111	171

Class activity

		Actual		
		None	Short	Long
Predicted	None	210	111	170
	Short	210	111	170
	Long	209	111	171

> What is the accuracy of our predictions in this confusion matrix?

Class activity

> What would our confusion matrix look like if for every individual, we just predicted the most common contraception choice in the data?

Class activity

> What would our confusion matrix look like if for every individual, we just predicted the most common contraception choice in the data?

The most common choice is None, so

		Actual		
		None	Short	Long
Predicted	None	629	333	511
	Short	0	0	0
	Long	0	0	0

Class activity

		Actual		
		None	Short	Long
Predicted	None	629	333	511
	Short	0	0	0
	Long	0	0	0

> What is the accuracy of our predictions in this confusion matrix?

Class activity

Do we do better than random guessing?

Moral

- By itself, accuracy isn't particularly useful for summarizing prediction performance
- We need to look at predictive ability for each class