# SOC542 Statistical Methods in Sociology II

## Binary outcomes II

Thomas Davidson

Rutgers University

March 28, 2022

## Course updates

- Homework 3 is due 4/1
- Start working on replication projects as soon as possible
    - Begin by replicating one of the main findings

# Plan

- Interaction terms
- Predictions
- Marginal effects

# Logistic regression refresher

### Binary outcomes and logistic regression

- We are continuing to consider binary outcome variables, focusing on logistic regression:

$$p_i = logit^{-1}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} + ... + \beta_k x_{ki})$$

$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} + ... + \beta_k x_{ki})}}$$

- The goal is to estimate $p_i$, the probability that the outcome $y = 1$ as a function of covariates.
- Logistic regression is a generalized linear model, where a link function is used to project a linear model onto a non-linear outcome.

# Logistic regression refresher

### Binary outcomes and logistic regression

- The $\beta$ coefficients in a logistic regression are *log-odds*.
- *exp*$(\beta)$ can allows us to interpret these coefficients as *odds-ratios*.
- We can make predictions to obtain *probabilities*.
  - $\beta_x/4$ provides an upper-bound for the effect of a unit-change in $x$ on $p_i$.

# Interaction terms

### Specifying an interaction

- If we expect there to be an **interaction** between $x$ and $z$, such that the effect of $x$ on $y$ varies according to the level of $z$, we can add an **interaction term** into our model formula.

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + u$$

- $\beta_0$ and $\beta_1$ are now considered as the **main effects**.
- $\beta_3$ is the coefficient for the interaction term, representing the effect of $x$ times $z$.

# Interaction terms

### Specifying an interaction

- If we're estimating an LPM we can use the same formula as above.
- For a logistic regression, we specify an interaction in the same way within the link function:

$$P(y = 1) = p = logit^{-1}(\beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz)$$

# Data

### Diffusion of Microfinance[1]

- Survey data from 75 villages in Karnataka, India
  - Focus only on women aged 18-65 and 72 villages
  - Listwise deletion used to drop respondents missing key variables
  - N = 8976
- Dependent variable:
  - Membership in a micro-finance Self-Help Group (SHG), N = 3357
- Independent variables:
  - Age (continuous)
  - Nativity (dummy)
    - 72% of women not born in current village, largely due to marriage-related migration

[1] Data from Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341 (6144): 1236498–1236498. Link to paper. Harvard Dataverse link

# Interaction terms

### Data exploration

There are two different factors that will be useful for understanding the results. First, nonnative respondents (typically married women due to village exogamy) and SHG participants tend to be older than natives and non-participants.

```
data %>% group_by(nonnative) %>% summarize(mean(age), median(age))
```

```
## # A tibble: 2 x 3
##   nonnative `mean(age)` `median(age)`
##       <dbl>       <dbl>         <dbl>
## 1         0        30.5            28
## 2         1        36.7            35
```

# Interaction terms

### Data exploration

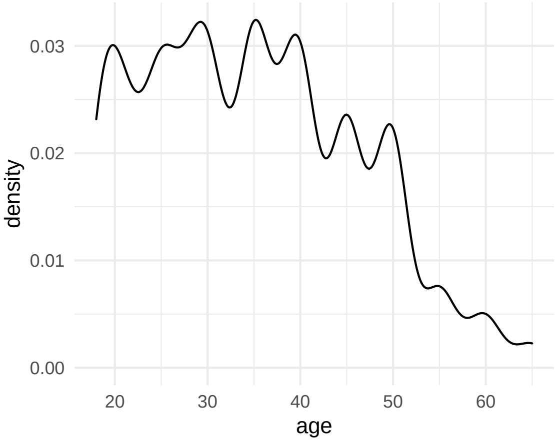
There are two different factors that will be useful for understanding the results. First, nonnative respondents (typically married women due to village exogamy) and SHG participants tend to be older than natives and non-participants.

```
data %>% group_by(shg) %>% summarize(mean(age), median(age))
```

```
## # A tibble: 2 x 3
##     shg `mean(age)` `median(age)`
##   <dbl>       <dbl>         <dbl>
## 1     0        34.1            32
## 2     1        36.4            35
```

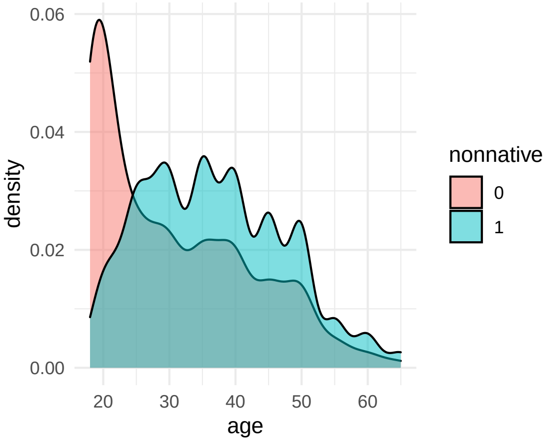
# Interaction terms

## Data exploration



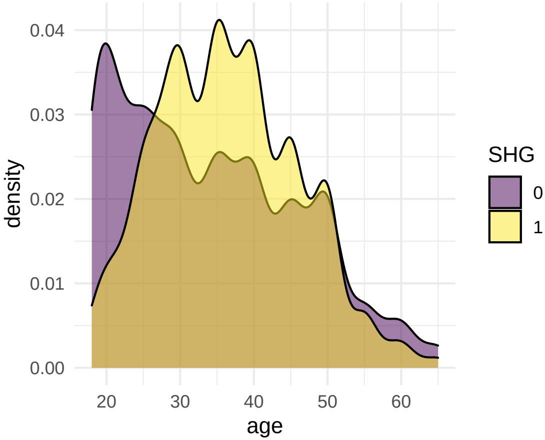
# Interaction terms

## Data exploration



# Interaction terms

## Data exploration



# Interaction terms

### Data exploration

Second, ~40% of nonnative women participate in SHGs, compared to only ~30% of natives.

```
data %>% group_by(nonnative, shg) %>%
    summarize(count = n(), .groups = "keep") %>% kable()
```

| nonnative | shg | count |
|---|---|---|
| 0 | 0 | 1768 |
| 0 | 1 | 751 |
| 1 | 0 | 3865 |
| 1 | 1 | 2592 |

# Interaction terms

### Estimating models

A LPM and logistic regression are used to estimate the probability of SHG membership as a function of age and nativity (whether a respondent was born in their current village of residence). We'll ignore any village fixed-effects to keep things simple.

```
lpm <- lm(shg ~ age + nonnative + age:nonnative,
          data = data)
logistic <- glm(shg ~ age + nonnative + age:nonnative,
                data = data, family = binomial())
```

# Interaction terms

## Comparing models

| | LPM | Logistic |
|---|---|---|
| (Intercept) | 0.024 | -2.184*** |
| | (0.027) | (0.130) |
| age | 0.009*** | 0.042*** |
| | (0.001) | (0.004) |
| nonnative | 0.337*** | 1.615*** |
| | (0.034) | (0.159) |
| age × nonnative | -0.008*** | -0.037*** |
| | (0.001) | (0.004) |
| Num.Obs. | 8976 | 8976 |
| R2 | 0.023 | |
| R2 Adj. | 0.022 | |
| Log.Lik. | -6109.447 | -5818.085 |
| F | 69.564 | 64.991 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# Interaction terms

### Intepretations

- In both models, the coefficients for the main effects of age and nativity are positive.
- The coefficients for interaction terms are both negative.
  - This implies that there is a negative effect of age for nonnative women. In other words, as age increases the probability of belonging to an SHG decreases.
- However, it is difficult to make sense of these interactions by only considering the coefficients, since the relationship between variables in a logistic regression is non-linear.

# Predictions

### Understanding interactions using predictions

- One of the ways we can start to make sense of these interactions is by making predictions.
- Let's consider predictions for a nonnative woman aged 25:

```
c1 <- coefficients(lpm)
c2 <- coefficients(logistic)
p.lpm <- as.numeric(c1[1] + c1[2]*25 + c1[3] + c1[4]*25)
print(p.lpm)
```

```
## [1] 0.3885258
```

```
p.logit <- invlogit(as.numeric(c2[1] + c2[2]*25 + c2[3] + c2[4]*25))
print(p.logit)
```

```
## [1] 0.3885585
```

# Predictions

### Understanding interactions using predictions

- The predictions are quite different if we ignore the interaction term:

```
p.lpm.ignore <- as.numeric(c1[1] + c1[2]*25 + c1[3])
p.logit.ignore <- invlogit(as.numeric(c2[1] + c2[2]*25 + c2[3]))
print(p.lpm.ignore)
```

```
## [1] 0.5855933
```

```
print(p.logit.ignore)
```

```
## [1] 0.6184885
```

# Predictions

### Understanding interactions using predictions

- We could also make the same predictions for native women, holding age constant.
- The equation is simplied since the main effect and interaction effect are now zero:

```
p.lpm2 <- as.numeric(c1[1] + c1[2]*25)
p.logit2 <- invlogit(as.numeric(c2[1] + c2[2]*25))
print(p.lpm2)
```

```
## [1] 0.2483897
```

```
print(p.logit2)
```

```
## [1] 0.2437525
```

- Despite the negative interaction, the main effect of nativity implies that a village native will be less likely to belong to an SHG, holding age constant.

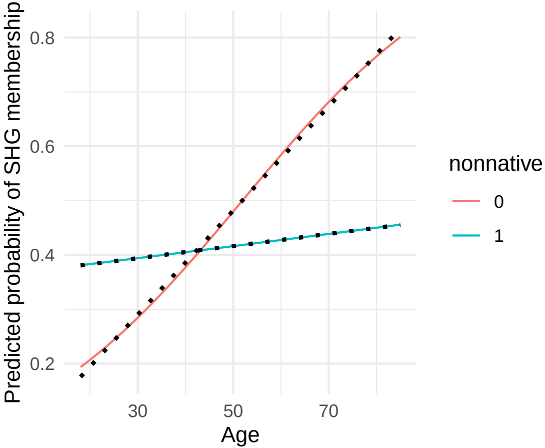
# Predictions

### Using a for-loop to make predictions

```
results <- as_tibble()
for (a in 18:85) {
    for (n in 0:1) {
        p <- invlogit(
            as.numeric(c2[1] + c2[2]*a + c2[3]*n +
                                            c2[4]*(a*n))
            )
        r <- list("p" = p, "age" =  a, "nonnative" = n)
        results <- bind_rows(results, r)
    }
}
```

# Predictions

## Predicted values by age and nativity




Dotted black lines show linear fits to the predictions to help illustrate non-linearity. Neither line is truly linear, although the characteristic S-curve is only discernable for the steeper line.

# Predictions

## Using the `predictions` function

- These predictions can be obtained by using the `predictions` function from `marginaleffects`.[2]

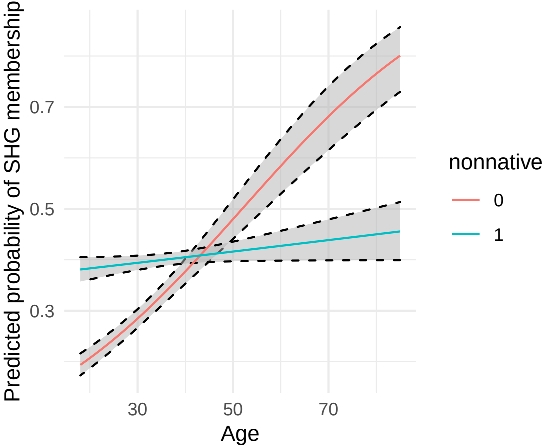
```
new <- results %>% select(age, nonnative)
preds <- predictions(logistic, newdata = as.data.frame(new))
preds %>% select(predicted, std.error, age, nonnative) %>%
    head(5) %>% kable()
```

| predicted | std.error | age | nonnative |
|---:|---:|---:|---:|
| 0.194 | 0.011 | 18 | 0 |
| 0.381 | 0.012 | 18 | 1 |
| 0.200 | 0.011 | 19 | 0 |
| 0.382 | 0.012 | 19 | 1 |
| 0.207 | 0.011 | 20 | 0 |

[2] Standard errors are calculated using an approach known as the delta method. See this post for further details.

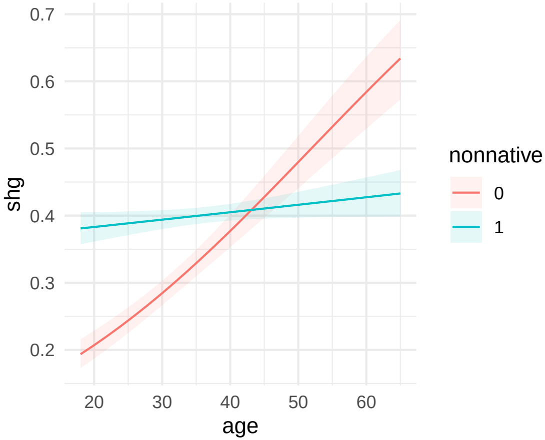
# Predictions

## Plotting the results



# Predictions

We can directly obtain these results by using the `plot_cap` function, where CAP stands for "Conditional Adjusted Predictions".



# Predictions

### Improving the model

- The previous model suggests quite different patterns by group.
  - For natives, there is a strong positive relationship between age and SHG membership.
  - For nonnatives, there is little evidence of such a relationship.
- Although there are age differences, these patterns seem remarkably strong.
  - I suspect that adding a squared term for age will help to improve the model.

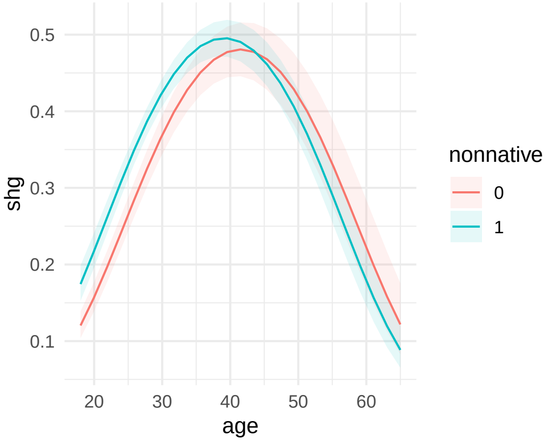
# Predictions

| | Logistic 1 | Logistic 2 |
|---|---|---|
| (Intercept) | -2.184*** | -6.030*** |
| | (0.130) | (0.261) |
| age | 0.042*** | 0.287*** |
| | (0.004) | (0.014) |
| nonnative | 1.615*** | 0.737*** |
| | (0.159) | (0.181) |
| age $\times$ nonnative | -0.037*** | -0.017*** |
| | (0.004) | (0.005) |
| I(age^2) | | -0.003*** |
| | | (0.000) |
| Num.Obs. | 8976 | 8976 |
| Log.Lik. | -5818.085 | -5638.019 |
| F | 64.991 | 121.010 |

$+$ p $<$ 0.1, * p $<$ 0.05, ** p $<$ 0.01, *** p $<$ 0.001

# Predictions

## Making new predictions



# Marginal effects

### Predictions versus marginal effects

- Predictions and associated plots allow us to observe differences on the outcome scale (in this case probabilities) across different values of the data.
- But what if we want to make statements about the overall effect of a predictor?
  - What is the average effect of age?
  - How does the effect of age vary as a function of other covariates?
- Like polynomial regression, it is difficult to determine this by examining coefficients or plotting predictions.

# Marginal effects

## Definitions

- A **marginal effect** is the relationship between change in single predictor and the dependent variable while *holding other variables constant*.
  - Recall that standard OLS coefficients can be intepreted as marginal effects, but this is no longer true with logistic regression.
- "Marginal effects are partial derivatives of the regression equation with respect to each variable in the model for each unit in the data."[3]
- In a logistic regression, the effects of predictors can be non-linear, such that the effect of $x$ on $y$ will vary according to the level of $z$.

[3] Leeper 2021. See the vignette for the margins package.

# Marginal effects

### Calculation

- We have previously used the `margins` package to estimate marginal effects.
- All analyses in this lecture use the `marginaleffects` package, which extends the functionality of `margins` and works for both frequentist and Bayesian models.[4]

[4] Read the documentation provided here for further information.

# Marginal effects

### Calculation

Observe how the `marginaleffects` function returns $N * k$ rows, where $N$ is the number of observations in the dataset and $k$ is the number of *unique* predictors.

```
ME <- marginaleffects(logistic2)
dim(ME)
```

```
## [1] 17952     9
```

```
dim(data)[1]*2
```

```
## [1] 17952
```

# Marginal effects

### Interpretation

This table shows the marginal effects for age and nonnative for the first two respondents.

```
ME %>% filter(rowid <= 2) %>% arrange(desc(rowid)) %>%
    select(rowid, term, dydx, std.error, shg, age, age.1, nonnative) %>%
    kable()
```

| rowid | term | dydx | std.error | shg | age | age.1 | nonnative |
|---:|---|---:|---:|---:|---:|---:|---:|
| 2 | age | 0.022 | 0.001 | 1 | 24 | 576 | 1 |
| 2 | nonnative | 0.071 | 0.017 | 1 | 24 | 576 | 1 |
| 1 | age | 0.020 | 0.001 | 0 | 27 | 729 | 1 |
| 1 | nonnative | 0.066 | 0.016 | 0 | 27 | 729 | 1 |

# Marginal effects

### Marginal effects at specified values

- Marginal effects are better understood by contextualizing them at relevant values of the data.
- Like the example above, we may want to calculate the marginal effect of a predictor at specific values of other covariates.
  - e.g. What is the marginal effect of nativity for women aged 25?
  - e.g. What is the marginal effect of age for nonnative women?

# Marginal effects

## Marginal effects at specified values: Nativity for age 25

```
ME.n <- marginaleffects(logistic2,
                        newdata = datagrid(nonnative = c(0,1),
                                           age = c(25)))
ME.n %>% filter(term == "nonnative") %>%
    select(dydx, std.error, nonnative, age) %>%
    head() %>% kable()
```

| dydx | std.error | nonnative | age |
|---|---|---|---|
| 0.062 | 0.013 | 0 | 25 |
| 0.070 | 0.017 | 1 | 25 |

# Marginal effects

### Marginal effects at specified values: Comparing 25 and 65 year olds

- We can use these values to calculate the difference by nativity:

```
round(ME.n$dydx[4] - ME.n$dydx[3],3)
```

```
## [1] 0.008
```

- Here is the same result when considering respondents aged 65 (calculation omitted):

```
## [1] -0.011
```

# Marginal effects

## Marginal effects at specified values: Age for natives

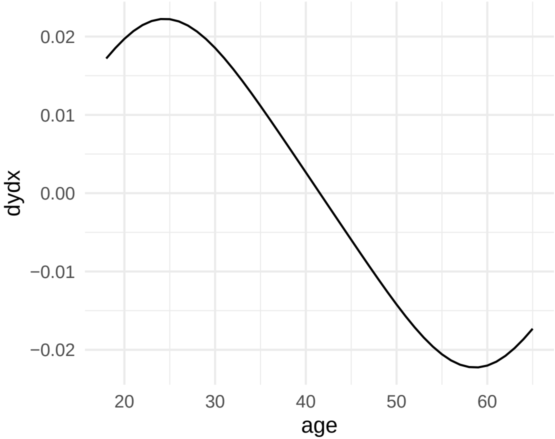
```
ME.a <- marginaleffects(logistic2,
                        newdata = datagrid(nonnative = 0,
                                          age = 18:65))
ME.a %>% filter(term == "age") %>%
    select(dydx, std.error, nonnative, age)  %>%
    head() %>% kable()
```

| dydx | std.error | nonnative | age |
|---:|---:|---:|---:|
| 0.017 | 0.001 | 0 | 18 |
| 0.019 | 0.001 | 0 | 19 |
| 0.020 | 0.001 | 0 | 20 |
| 0.021 | 0.001 | 0 | 21 |
| 0.021 | 0.001 | 0 | 22 |
| 0.022 | 0.001 | 0 | 23 |

# Marginal effects

## Marginal effects at specified values: Age for natives

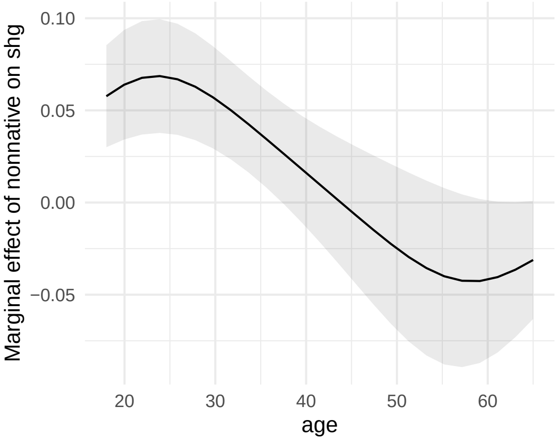
In this case we have a marginal effect for every value of age, so it makes sense to create a plot.



# Marginal effects

## Plotting conditional marginal effects using `plot_cme`

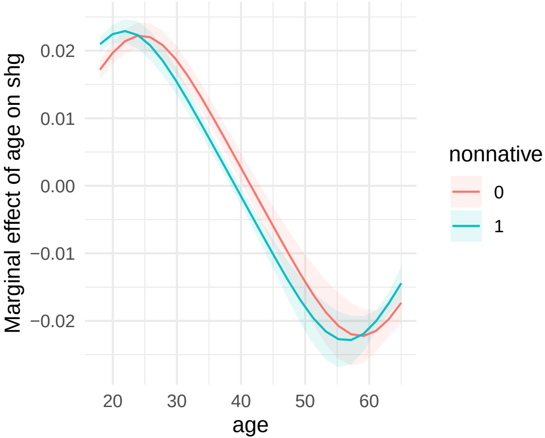
```
plot_cme(logistic2, effect = "nonnative", condition = c("age"))
```



# Marginal effects

## Plotting conditional marginal effects using `plot_cme`

```
plot_cme(logistic2, effect = "age", condition = c("age", "nonnative"))
```



# Marginal effects

### Marginal effects at means

- A common approach is to assess the **marginal effects at means (MEM)**, examining the marginal effect of change in a predictor while holding other covariates at their average values.
- This can be convenient if we don't have any clear reasons for selecting particular values to examine.

# Marginal effects

### Marginal effects at means

By default, we get the MEM if we specify an empty data grid. However, the mean for nonnative doesn't really make sense.

```
marginaleffects(logistic2, newdata = datagrid()) %>%
    kable()
```

| rowid | type | term | dydx | std.error | age | nonnative |
|---:|---|---|---:|---:|---:|---:|
| 1 | response | age | 0.008 | 0.001 | 34.951 | 0.719 |
| 1 | response | nonnative | 0.037 | 0.013 | 34.951 | 0.719 |

# Marginal effects

## Marginal effects at means

In this case, it is more appropriate to consider the marginal effects for each value of nonnative. If we're just interested in the modal category, we could consider the rows where `nonnative = 1`.

```
marginaleffects(logistic2,
                newdata = datagrid(nonnative = c(0,1))) %>%
                kable()
```

| rowid | type | term | dydx | std.error | age | nonnative |
|---:|---|---|---|---:|---:|---:|
| 1 | response | age | 0.011 | 0.001 | 34.951 | 0 |
| 2 | response | age | 0.007 | 0.001 | 34.951 | 1 |
| 1 | response | nonnative | 0.037 | 0.013 | 34.951 | 0 |
| 2 | response | nonnative | 0.037 | 0.013 | 34.951 | 1 |

# Marginal effects

### Average marginal effects

- A different approach involves averaging over the variation in other covariates to calculate the **average marginal effect (AME)** of a predictor.
- We can obtain this by averaging over all the observation specific marginal effects.

# Marginal effects

### Average marginal effects

We can obtain the AME by taking a summary of the marginal effects table produced above (`ME`).

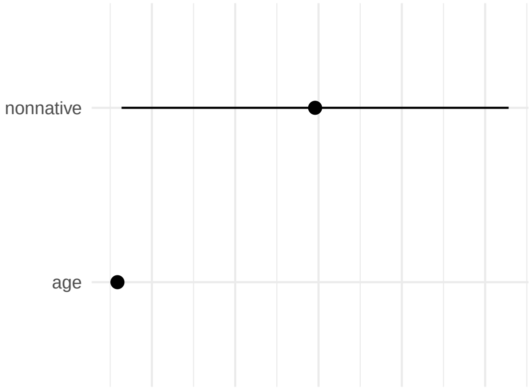
```
AME <- summary(ME)
AME %>% kable()
```

| type | term | estimate | std.error | statistic | p.value | conf.low | conf. |
|---|---|---|---|---|---|---|---|
| response | age | 0.006 | 0.000 | 16.159 | 0.000 | 0.005 | 0 |
| response | nonnative | 0.030 | 0.012 | 2.497 | 0.013 | 0.006 | 0 |

# Marginal effects

We can also produce a plot of the marginal effects and associated confidence intervals by calling `modelplot` on the full marginal effects table.

```
modelplot(ME)
```



# Marginal effects

### Improving the model?

- These models show how the marginal effect of age is highly non-linear
- Perhaps an additional polynomial for age would further improve the fit

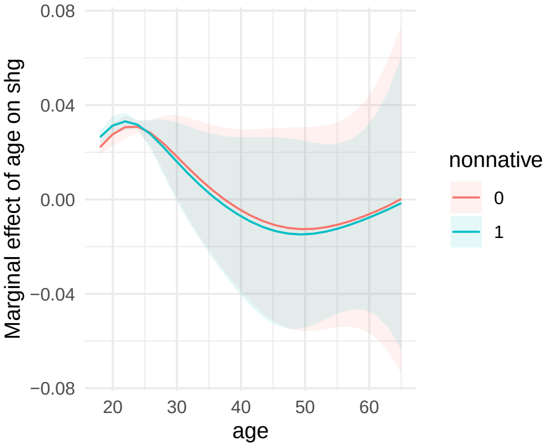
# Marginal effects

| | Logistic 1 | Logistic 2 | Logistic 3 |
|---|---|---|---|
| (Intercept) | -2.184*** | -6.030*** | -10.651*** |
| | (0.130) | (0.261) | (0.745) |
| age | 0.042*** | 0.287*** | 0.699*** |
| | (0.004) | (0.014) | (0.063) |
| nonnative | 1.615*** | 0.737*** | 0.457* |
| | (0.159) | (0.181) | (0.187) |
| age $\times$ nonnative | -0.037*** | -0.017*** | -0.010* |
| | (0.004) | (0.005) | (0.005) |
| I(age^2) | | -0.003*** | -0.015*** |
| | | (0.000) | (0.002) |
| I(age^3) | | | 0.000*** |
| | | | (0.000) |
| Log.Lik. | -5818.085 | -5638.019 | -5615.637 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# Marginal effects

```
plot_cme(logistic3, effect = "age", condition = c("age", "nonnative"))
```



# Marginal effects

### Bayesian estimation

- The same approaches apply to Bayesian models. The only difference is that the uncertainity in the posterior distribution must be incorporated into the calculation of the marginal effects.
- Fortunately for us, the `marginaleffects` package can handle models estimated using `rstanarm`.

# Marginal effects

### Bayesian estimation

```
bayes <- stan_glm(shg ~ age + I(age^2) + nonnative + age:nonnative,
                  data = data, family = binomial(),
                  chains = 1, refresh = 0)
```

# Marginal effects

## Bayesian estimation

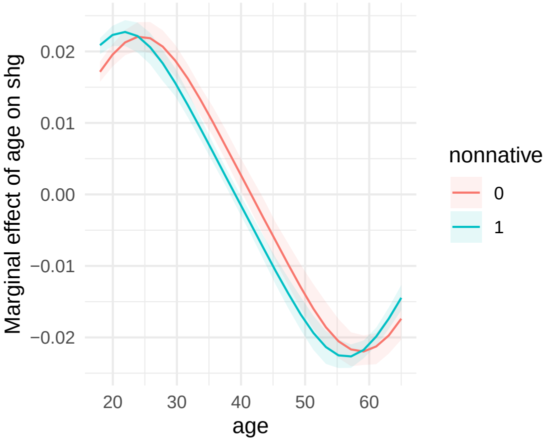
The AMEs are close to those obtained from the maximum likelihood model.

```
summary(marginaleffects(bayes)) %>%
    kable()
```

| type | term | contrast | estimate | conf.low | conf.high |
|---|---|---|---|---|---|
| response | age | dydx | 0.006 | 0.004 | 0.008 |
| response | nonnative | dydx | 0.030 | -0.002 | 0.061 |

# Marginal effects

We can see similar relationships using the same `plot_cme` specification as above.



# Marginal effects

### Improving the model?

```
bayes.2 <- stan_glm(shg ~ age + I(age^2) + I(age^3) + nonnative + age:n
                    data = data, family = binomial(),
                    chains = 1,  refresh = 0)
```

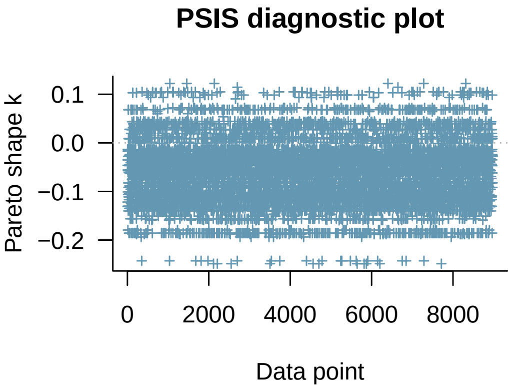
# Marginal effects

```
l1 <- loo(bayes)
l2 <- loo(bayes.2)
loo_compare(l1,l2)
```

```
##          elpd_diff se_diff
## bayes.2   0.0       0.0
## bayes   -17.8       3.9
```

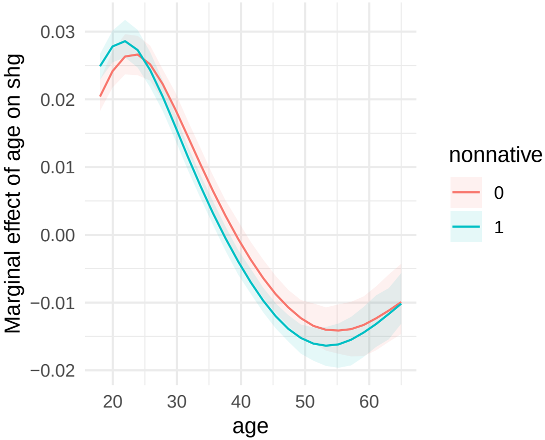
# Marginal effects

```
plot(l2)
```



# Marginal effects

## Improving the model?



# Marginal effects

### Comparing the Bayesian and Maximum Likelihood approaches

- Both approaches to estimation produce substantively similar results.
- The Bayesian approach appears to produce more stable predictions for more complex parameterizations.
  - In both cases, the model with $age^3$ appears to improve fit, but the marginal effects plots are very noisy for the MLE approach.

## Summary

- Logistic regression models (and other GLMs) can be challenging to interpret, particularly when we add interaction terms.
- By making predictions, we can observe variation in outcomes across different covariate values and make interpretations on the probability scale.
- Marginal effects allow us to better isolate the effect of individual variables, akin to the way we interpret OLS results.
- In both cases, visualizations help us to better understand the interactions between key variables.

# Next week

- Count outcomes
  - Poisson regression
  - Negative-binomial regression
  - And zero-inflated variants