

§2.6 Interpreting parameters

The reason we fit models in the first place is because the fitted parameters can (hopefully) tell us something useful about our data.

To understand what they tell us, we need to know how to interpret them.

Exercise 2.6.2 (Contrasts)

In the dataset below, of measurements from two groups A and B, interpret the parameters from these models:

$$y \approx \alpha 1_{g=A} + \beta 1_{g=B} \quad (\text{M1})$$

$$y \approx \alpha' + \beta' 1_{g=B} \quad (\text{M2})$$

$$y \approx \alpha'' + \beta'' 1_{g=A} + \gamma'' 1_{g=B} \quad (\text{M3})$$

g	y	1	$1_{g=A}$	$1_{g=B}$
A	0.5	1	1	0
A	1.9	1	1	0
B	3.5	1	0	1
B	1.1	1	0	1
B	2.3	1	0	1

These three models all have the same feature space

M1

$$y \approx \alpha$$

$$y \approx \alpha$$

$$y \approx \beta$$

$$y \approx \beta$$

$$y \approx \beta$$

M1 tells us the average response in each group

M2

$$y \approx \alpha'$$

$$y \approx \alpha' + \beta'$$

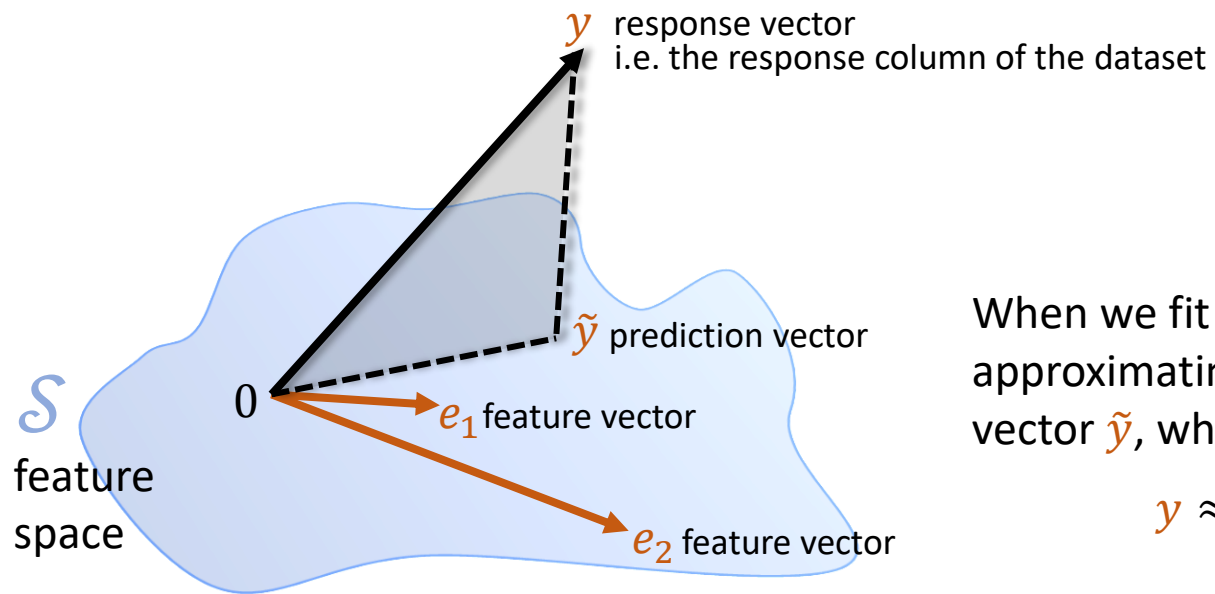
M2 tells us β' , the difference between the two groups

M3

$$y \approx \alpha'' + \beta''$$

$$y \approx \alpha'' + \gamma''$$

The parameters in M3 aren't pinned down because its features are linearly dependent.



When we fit a linear model, we are in effect approximating the response vector y by a prediction vector $\tilde{y}$, which is a point in the feature space:

$$y \approx \beta_1 e_1 + \beta_2 e_2 \in \text{span}(\{e_1, e_2\})$$

The same prediction vector $\tilde{y}$ may be written in different ways, depending on the feature space we're using.

models with
features $1_{g=A}, 1_{g=B}$

$$0.3 \, 1_{g=A} + 1.2 \, 1_{g=B}$$

$$2.3 \, 1_{g=A}$$

$$0.6 \, 1_{g=A} - 1.5 \, 1_{g=B}$$

models with
features $1, 1_{g=B}$

$$0.3 \, 1 + 0.9 \, 1_{g=B}$$

models with
features $1, 1_{g=A}, 1_{g=B}$

$$0.1 \, 1 + 0.2 \, 1_{g=A} + 1.1 \, 1_{g=B}$$

$$0.3 \, 1 + 0.9 \, 1_{g=B}$$

$$0.3 \, 1_{g=A} + 1.2 \, 1_{g=B}$$

- Write out the predicted response for a few typical / representative datapoints.
This helps see what the parameters report.
- It may be useful to use different features (with the same feature space).
This doesn't change the model's predictions, but it does change what the parameters report.
- Check if the features are linearly dependent.
If so, we say the parameters are *non-identifiable*. We may get different answers each time we fit the model.

 Sign in



Stop and search

🕒 This article is more than 3 years old

Met police 'disproportionately' use stop and search powers on black people

**London's minority black population
targeted more than white population in
2018 - official figures**

The Guardian

News website of the year

Do I trust this finding?

- ❖ Proportionate to what?
- ❖ Is it cherry picking?
- ❖ Why the scare quotes?

 Sign in



Stop and search

🕒 This article is more than 3 years old

Met police 'disproportionately' use stop and search powers on black people

**London's minority black population
targeted more than white population in
2018 - official figures**

The
Guardian
News website of the year

Can I set up a model with
a parameter that
measures the quantity
I'm interested in?

force	Date	LatLng	Object of search	Gender	Age range	Officer-defined ethnicity	Outcome
cambridgeshire	2023-08-31 15:44:04+00:00	(52.43,-0.142)	Controlled drugs				A no further action disposal
cambridgeshire	2023-08-31 15:35:41+00:00	(52.43,-0.142)	Firearms	Male	25-34	White	Khat or Cannabis warning
cambridgeshire	2023-08-31 14:44:04+00:00	(52.43,-0.142)	Firearms	Male	25-34	White	Khat or Cannabis warning
cambridgeshire	2023-08-31 03:44:14+00:00	(52.58,-0.244)	Offensive weapons	Male		Other	A no further action disposal
cambridgeshire	2023-08-31 02:34:16+00:00	(52.59,-0.247)	Controlled drugs	Male	25-34	White	Arrest
cambridgeshire	2023-08-31 02:27:10+00:00	(52.21,0.124)	Controlled drugs	Male	18-24	White	A no further action disposal
cambridgeshire	2023-08-30 22:28:13+00:00	(52.45,-0.117)	Controlled drugs	Female	over 34	White	A no further action disposal
cambridgeshire	2023-08-30 20:24:13+00:00	(52.32,-0.0708)	Controlled drugs	Male	10-17	White	Summons / charged by post
cambridgeshire	2023-08-30 14:26:58+00:00	(52.57,-0.24)	Controlled drugs	Male	over 34	Asian	A no further action disposal
cambridgeshire	2023-08-30 14:13:45+00:00	(52.57,-0.24)	Controlled drugs	Male	25-34	Black	Arrest

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y \approx \alpha + \beta_{\text{eth}}$$

where eth_i is the officer-defined ethnicity for record i , and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

- a) Write this as a linear model using one-hot coding.
- b) Are the parameters identifiable? If not, rewrite the model so that they are.
- c) Does the model suggest there is racial bias in policing actions?

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y \approx \alpha + \beta_{\text{eth}}$$

where eth_i is the officer-defined ethnicity for record i , and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

Write out some predicted responses.

For people with $\text{eth} = \text{Asian}$, $y \approx \alpha + \beta_{\text{Asian}}$

For people with $\text{eth} = \text{Mixed}$, $y \approx \alpha + \beta_{\text{Mixed}}$
etc.

} the prediction formula says roughly what y is likely to be

We can also write the average response in ethnic group k with simple counting:

$$\text{avg. } y \text{ in group } k = \frac{\sum_{i:\text{eth}_i=k} y_i}{\#\text{stops with eth } k} = \frac{\#\text{finds with eth } k}{\#\text{stops with eth } k} = \mathbb{P}(\text{find something} \mid \text{eth} = k)$$

Call this the "hit rate"

This model is thus proposing: $\mathbb{P}(\text{find something}) \approx \alpha + \beta_k$ for a person in ethnic group k

If $\beta_k < 0$, that means $\mathbb{P}(\text{find something})$ is low compared to other ethnic groups, i.e. the police are stopping relatively more innocent people.

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y \approx \alpha + \beta_{\text{eth}}$$

where eth_i is the officer-defined ethnicity for record i , and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

a) Write this as a linear model using one-hot coding.

eth has five levels:

Asian
Black
Mixed
Other
White

$$\vec{y} = \alpha \vec{1} + \beta_{As} 1_{\vec{eth}=As} + \beta_{Bl} 1_{\vec{eth}=Bl} + \beta_{Mix} 1_{\vec{eth}=Mix} + \beta_{oth} 1_{\vec{eth}=oth} + \beta_{wh} 1_{\vec{eth}=wh}$$

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y \approx \alpha + \beta_{\text{eth}}$$

where eth_i is the officer-defined ethnicity for record i , and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

a) Write this as a linear model using one-hot coding.

$$y \approx \alpha \mathbf{1} + \underset{\text{Asian}}{\beta_{\text{As}}} \mathbf{1}_{\text{eth}=\text{As}} + \underset{\text{Black}}{\beta_{\text{Bl}}} \mathbf{1}_{\text{eth}=\text{Bl}} + \underset{\text{Mixed}}{\beta_{\text{Mi}}} \mathbf{1}_{\text{eth}=\text{Mi}} + \underset{\text{Other}}{\beta_{\text{Oth}}} \mathbf{1}_{\text{eth}=\text{Oth}} + \underset{\text{White}}{\beta_{\text{Wh}}} \mathbf{1}_{\text{eth}=\text{Wh}}$$

```
1 ethnicity_levels = np.unique(eth)
2 eth_onehot = [np.where(eth==k,1,0) for k in ethnicity_levels]
3
4 model = sklearn.linear_model.LinearRegression()
5 model.fit(np.column_stack(eth_onehot), y)
6 alpha, beta_s = model.intercept_, model.coef_
7
8 print(f'alpha = {alpha}')
9 for k, beta in zip(ethnicity_levels, beta_s):
10     print(f'beta[{k}] = {beta}')
```

```
alpha = -34037792910.00365
beta[Asian] = 34037792910.26522
beta[Black] = 34037792910.265717
beta[Mixed] = 34037792910.2939
beta[Other] = 34037792910.260
beta[White] = 34037792910.26
```



Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y \approx \alpha + \beta_{\text{eth}}$$

where eth_i is the officer-defined ethnicity for record i , and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

- a) Write this as a linear model using one-hot coding.
- b) Are the parameters identifiable? If not, rewrite the model so that they are.

(a) $y \approx \alpha \mathbf{1} + \beta_{\text{As}} \mathbf{1}_{\text{eth}=\text{As}} + \beta_{\text{Bl}} \mathbf{1}_{\text{eth}=\text{Bl}} + \beta_{\text{Mi}} \mathbf{1}_{\text{eth}=\text{Mi}} + \beta_{\text{Oth}} \mathbf{1}_{\text{eth}=\text{Oth}} + \beta_{\text{Wh}} \mathbf{1}_{\text{eth}=\text{Wh}}$

(b). No. The features are not linearly independent, thus the parameters are not identifiable.

We have a choice about how to rewrite the model:

$$\vec{y} = \beta_{\text{As}} \mathbf{1}_{\text{eth}=\text{As}} + \beta_{\text{Bl}} \mathbf{1}_{\text{eth}=\text{Bl}} + \dots$$

or $\vec{y} = \alpha \vec{\mathbf{1}} \quad \times \quad + \beta_{\text{Bl}} \mathbf{1}_{\text{eth}=\text{Bl}} + \dots$

or many other choices

Example 2.6.4

c) Does this rewritten model suggest there is racial bias in policing actions?

$$y \approx \alpha \mathbf{1} + \beta_{Bl} \mathbf{1}_{eth=Bl} + \beta_{Mi} \mathbf{1}_{eth=Mi} + \beta_{Oth} \mathbf{1}_{eth=Oth} + \beta_{Wh} \mathbf{1}_{eth=Wh}$$

First, let's interpret the parameters of the rewritten model.

For a person with $eth = As$ predicted $y = \alpha$

$$eth = Bl \quad \alpha + \beta_{Bl}$$

$$eth = Mi \quad \alpha + \beta_{Mi}$$

$$eth = Oth \quad \alpha + \beta_{Oth}$$

$$eth = Wh \quad \alpha + \beta_{Wh}$$

These β_k measure differences with respect to the baseline of people with $eth=Asian$.

If we had chosen a different rewriting, we'd get different parameter interpretation.
We should choose the features based on what quantities we want reported.

Example 2.6.4

c) Does this rewritten model suggest there is racial bias in policing actions?

$$y \approx \alpha \mathbf{1} + \beta_{\text{Bl}} \mathbf{1}_{\text{eth}=\text{Bl}} + \beta_{\text{Mi}} \mathbf{1}_{\text{eth}=\text{Mi}} + \beta_{\text{Oth}} \mathbf{1}_{\text{eth}=\text{Oth}} + \beta_{\text{Wh}} \mathbf{1}_{\text{eth}=\text{Wh}}$$

```
1 some_levels = [k for k in ethnicity_levels if k != 'Asian']
2 eth_onehot = [np.where(eth==k,1,0) for k in some_levels]
3
4 model = sklearn.linear_model.LinearRegression()
5 model.fit(np.column_stack(eth_onehot), y)
6
7 alpha, beta_s = model.intercept_, model.coef_
8
9 print(f'alpha = {alpha}')
10 for k, beta in zip(some_levels, beta_s):
11     print(f'beta[{k}] = {beta}')
```

```
alpha = 0.261423626892284
beta[Black] = 0.0008154705731562
beta[Mixed] = 0.028715617211154
beta[Other] = -0.00447105736658
beta[White] = -0.003720378247083
```



Are these meaningful?
Is this evidence of bias,
or is it just noise?
[See next two weeks]

Modelling makes it easy to ask sophisticated questions about our dataset

"hit rate"

$$y \approx \alpha + \beta_{\text{eth}}$$

the model we've been using

What does δ estimate?

$$y \approx \alpha + \beta_{\text{eth}} + \delta 1_{\text{force}=\text{Cambridgeshire}}$$

Assuming Cambridgeshire has the same ethnic bias as the rest of the country, is its base 'hit rate' the same as the rest of the country?

In other words, Cambridgeshire may have a different average when we look at crude raw counts. If so, is it explainable by the fact that Cambridgeshire has a different ethnic mix?

$$y \approx \alpha + \gamma_{\text{gender}}$$

$$y \approx \alpha + \beta_{\text{eth}} + \gamma_{\text{gender}}$$

$$y \approx \alpha + \beta_{\text{eth}} + \gamma_{\text{gender}} + \delta_{\text{eth} + \text{"_" + gender}}$$



```
In [1]: import numpy as np
import pandas
import matplotlib.pyplot as plt
import sklearn.linear_model
π = np.pi
```

Climate dataset challenge

- What is the rate of temperature increase in Cambridge?
- Are temperatures increasing at a constant rate, or has the increase accelerated?
- How do results compare across the whole of the UK?

Your task is to answer these questions using appropriate linear models, and to produce elegant plots to communicate your findings. Please submit a Jupyter notebook, or a pdf. Include explanations of what your models are, and of what your plots show.

The dataset is from <https://www.metoffice.gov.uk/pub/data/weather/uk/climate/>. Code for retrieving the dataset is given at the bottom.

Upload your answers
to Moodle by Sunday
evening
for presentation /
discussion next week

§2.4 Least squares estimation & probability



Carl Friedrich Gauss
1777–1855

Least squares estimation

Example 2.1.1

The Iris dataset has 50 records of iris measurements, from three species.

How does **Petal.Length (PL)** depend on **Sepal.Length (SL)**?

We fitted the linear model

$$PL \approx \alpha + \beta SL + \gamma SL^2$$

Fitting concept: find the prediction vector

$$\tilde{y} = \alpha + \beta SL + \gamma SL^2$$

in the feature space, that gets as close as possible to the response vector **PL**, according to the distance

$$\|PL - \tilde{y}\| = \sqrt{\sum_i (PL_i - \tilde{y}_i)^2}$$

Maximum likelihood estimation

Example

Let's fit the probability model

$$PL_i \sim \alpha + \beta SL_i + \gamma SL_i^2 + \text{Normal}(0, \sigma^2)$$

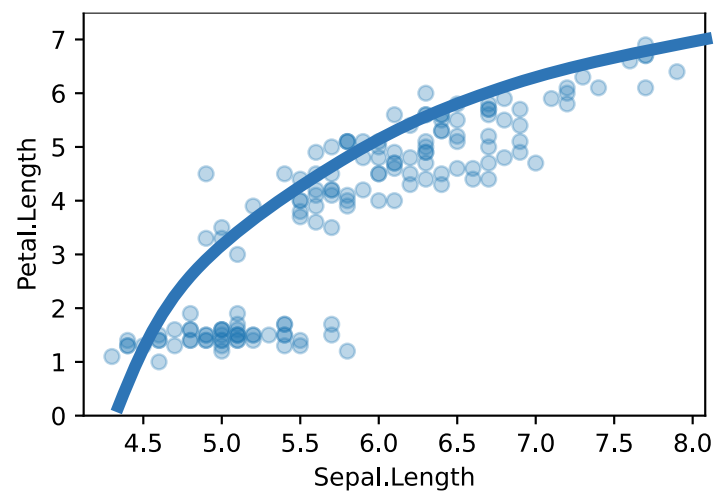
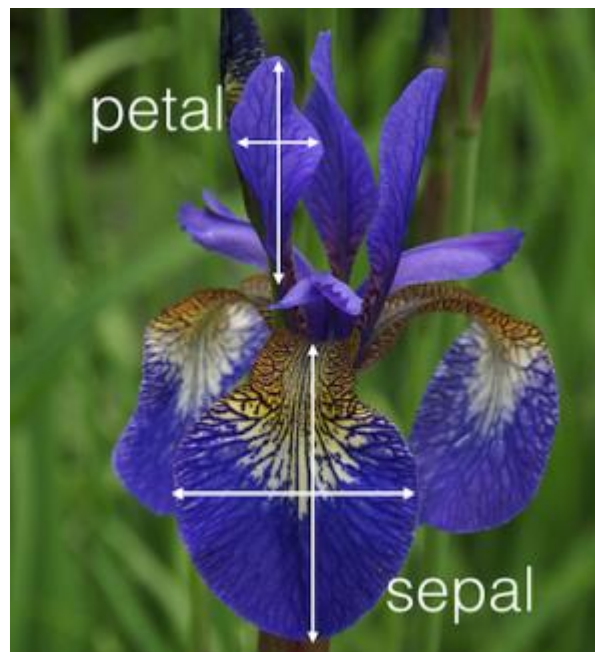
Fitting concept: maximum likelihood estimation

Example 2.1.1

The Iris dataset has 50 records of iris measurements, from three species.

Petal.Length	Petal.Width	Sepal.Length	Sepal.Width	Species
1.0	0.2	4.6	3.6	setosa
5.0	1.9	6.3	2.5	virginica
5.8	1.6	7.2	3.0	virginica
4.2	1.2	5.7	3.0	versicolor
...				

How does **Petal.Length** depend on **Sepal.Length**?



Let's guess that for parameters $\alpha, \beta, \gamma, \sigma$ (to be estimated),
Petal.Length $\approx \alpha + \beta$ **Sepal.Length** + γ (**Sepal.Length**)²

Model for a single observation: $PL_i \sim \alpha + \beta SL_i + \gamma SL_i^2 + N(0, \sigma^2)$

ie $Y_i \sim \alpha + \beta e_i + \gamma f_i + N(0, \sigma^2)$ relabeling, to save myself some writing:
 $y \equiv PL, e \equiv SL, f \equiv SL^2$

$\Rightarrow Y_i \sim N(\alpha + \beta e_i + \gamma f_i, \sigma^2)$ by linearity of the Gaussian

Likelihood of a single observation:

$$\Pr_{Y_i}(y_i; \alpha, \beta, \gamma, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[y_i - (\alpha + \beta e_i + \gamma f_i)]^2 / 2\sigma^2}$$

Log likelihood of the dataset:

$$\log \Pr(y_1 \dots y_n; \alpha, \beta, \gamma, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\alpha + \beta e_i + \gamma f_i)]^2$$

Maximum Likelihood Estimation:

We want to find $\alpha, \beta, \gamma, \sigma$ to maximize this

Maximize over the unknown parameters,
 α, β, γ , and σ :

$$\begin{aligned}
 & \max_{\alpha, \beta, \gamma, \sigma} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha + \beta e_i + \gamma f_i))^2 \right\} \\
 &= \max_{\sigma} \left[\max_{\alpha, \beta, \gamma} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - (\alpha + \beta e_i + \gamma f_i))^2 \right\} \right] \\
 &= \max_{\sigma} \left[-\frac{n}{2} \log(2\pi\sigma^2) + \max_{\alpha, \beta, \gamma} \left\{ -\frac{1}{2\sigma^2} \sum_i (y_i - (\alpha + \beta e_i + \gamma f_i))^2 \right\} \right] \\
 &= \max_{\sigma} \left[-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\{ \min_{\alpha, \beta, \gamma} \sum_i (y_i - (\alpha + \beta e_i + \gamma f_i))^2 \right\} \right] \\
 &= \max_{\sigma} \left[-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \hat{y}_i)^2 \right] \quad \text{where } \hat{y}_i = \hat{\alpha} + \hat{\beta} e_i + \hat{\gamma} f_i \\
 &\quad \text{obtained by least squares estimation} \\
 &\Rightarrow \hat{\sigma} = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}
 \end{aligned}$$

Maximize over the unknown parameters,
 α, β, γ , and σ :

Least squares estimation
derives from a Gaussian
 probability model.

If that model doesn't fit the
 data, then don't use least
 squares estimation!

$$\hat{\beta} e_i + \hat{\gamma} f_i$$

ained by
 squares estimation



Stop and search

🕒 This article is more than 3 years old

Met police 'disproportionately' use stop and search powers on black people

London's minority black population targeted more than white population in 2018 - official figures

Let $y_i \in \{0,1\}$ be the outcome for stop-and-search incident i .

What we did earlier:

$$y_i \approx \alpha + \beta_{\text{eth}_i} \quad \text{i.e. } Y_i \sim \alpha + \beta_{\text{eth}_i} + N(0, \sigma^2)$$

Fit α and $\beta_{\text{Bl}}, \beta_{\text{Mi}}, \dots$ using least squares estimation or, equivalently, fit using maximum likelihood estimation

What we should do:

$$Y_i \sim \text{Bin}(1, \alpha + \beta_{\text{eth}_i})$$

Fit the parameters using maximum likelihood estimation

though it'd be better to change the model slightly, to keep our coefficients are in the proper range, say $Y_i \sim \text{Bin}\left(1, \sigma(\alpha + \beta_{\text{eth}_i})\right)$ for some parameter transform $\sigma: \mathbb{R} \rightarrow [0,1]$