SOCI 620: Quantitative methods 2

Jan 26

Checking models and estimates

1. Interpretation with transformed variables
2. Prior predictive plots
3. Visualizing model predictions
4. Examining model fit
5. Visualizing predictions in R

Interpreting coefficients

$$\text{Income}_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{Age}_i$$

	Post. Mean
α	32813.3
β	185.7

If somebody ages from a_1 to a_2, how much do we expect their income to change?

$$\begin{aligned} \text{E}(I_i | A_i = a_2) - \text{E}((I_i | A_i = a_1) &= (\alpha + \beta a_2) - (\alpha + \beta a_1) \\ &= \beta(a_2 - a_1) \\ &= 185.7(a_2 - a_1) \end{aligned}$$

Interpreting coefficients

Predicting: **Income** ← Using: **Age**

$$\text{Income}_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{Age}_i$$

	Post. Mean
α	32813.3
β	185.7

Units of age | Years

Units of income | Dollars

Interpreting β | For each year of age, the model predicts about $186 more income per year.

Standardized variables

Predicting: **St(Income)** ⟵ Using: **St(Age)**

$$\text{St}(\text{Income}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

	Post. Mean
α	0
β	0.065

Units of age	Standard deviations of age
Units of income	Standard deviations of income
Interpreting β	For each standard deviation of age, the model predicts an increase of about 0.065 standard deviations in income.

Log of outcome variable

Predicting: **log(Inc)** ← Using: **Age**

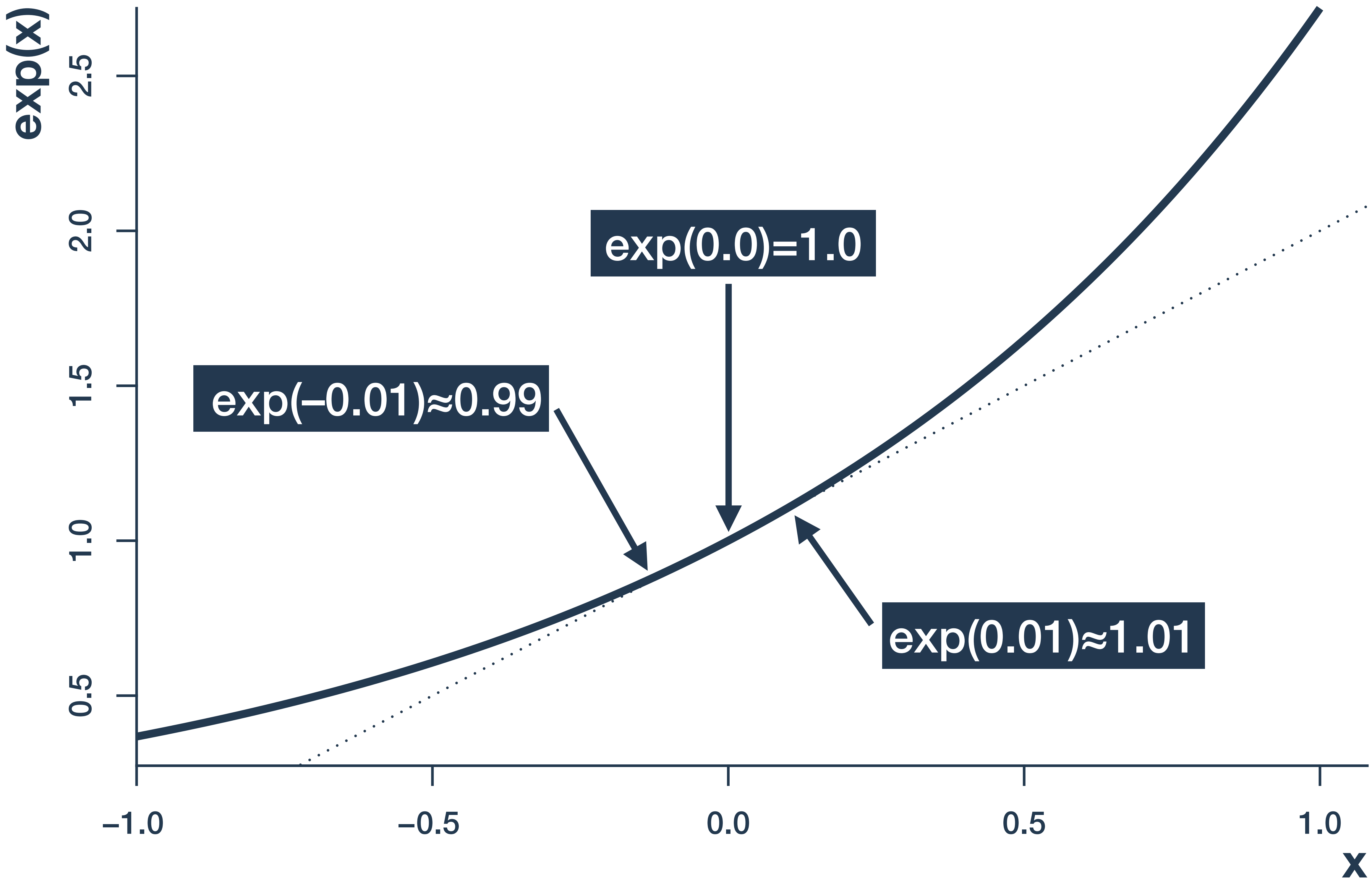
$$\log(\text{Income}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{Age}_i$$

	Post. Mean	exp(Mean)
α	9.648	15489.124
β	0.009	1.009

If somebody ages from a_1 to a_2 , how much do we expect their income to change?

$$\frac{\text{Inc}_2}{\text{Inc}_1} = \exp(\mathit{log}(\text{Inc}_2) - \log(\text{Inc}_1))$$
$$= \exp((\alpha + \beta a_2) - (\alpha + \beta a_1))$$
$$= \exp(\beta(a_2 - a_1))$$

Log of outcome variable



Log of outcome variable

Predicting: **log(Inc)** ← Using: **Age**

$$\log(\text{Income}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{Age}_i$$

	Post. Mean	exp(Mean)
α	9.648	15489.124
β	0.009	1.009

Units of age | Years

Units of income | Log dollars

Interpreting β | For each year of age, the model predicts a 0.9% increase in income.

Log of predictor variable

Predicting: **Income** ← Using: **log(Age)**

$$\text{Income}_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \log(\text{Age}_i)$$

	Post. Mean
α	–17586
β	15675

$$\begin{aligned} \text{Inc}_2 - \text{Inc}_1 &= (\alpha + \beta \log(a_2)) - (\alpha + \beta \log(a_1))) \\ &= \beta(\log(a_2) - \log(a_1))) \\ &= \beta(\log(a_2/a_1))) \end{aligned}$$

Log of predictor variable

Predicting: **Income** ← Using: **log(Age)**

$$\text{Income}_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \log(\text{Age}_i)$$

	Post. Mean
α	–17586
β	15675

Units of age	Log years
Units of income	Dollars
Interpreting β	For each 10% increase in age, the model predicts an increase of $\beta \times \log(1.1) = 15{,}675 \times 0.095 = 1{,}494.07$ dollars in income.

Prior predictive plot

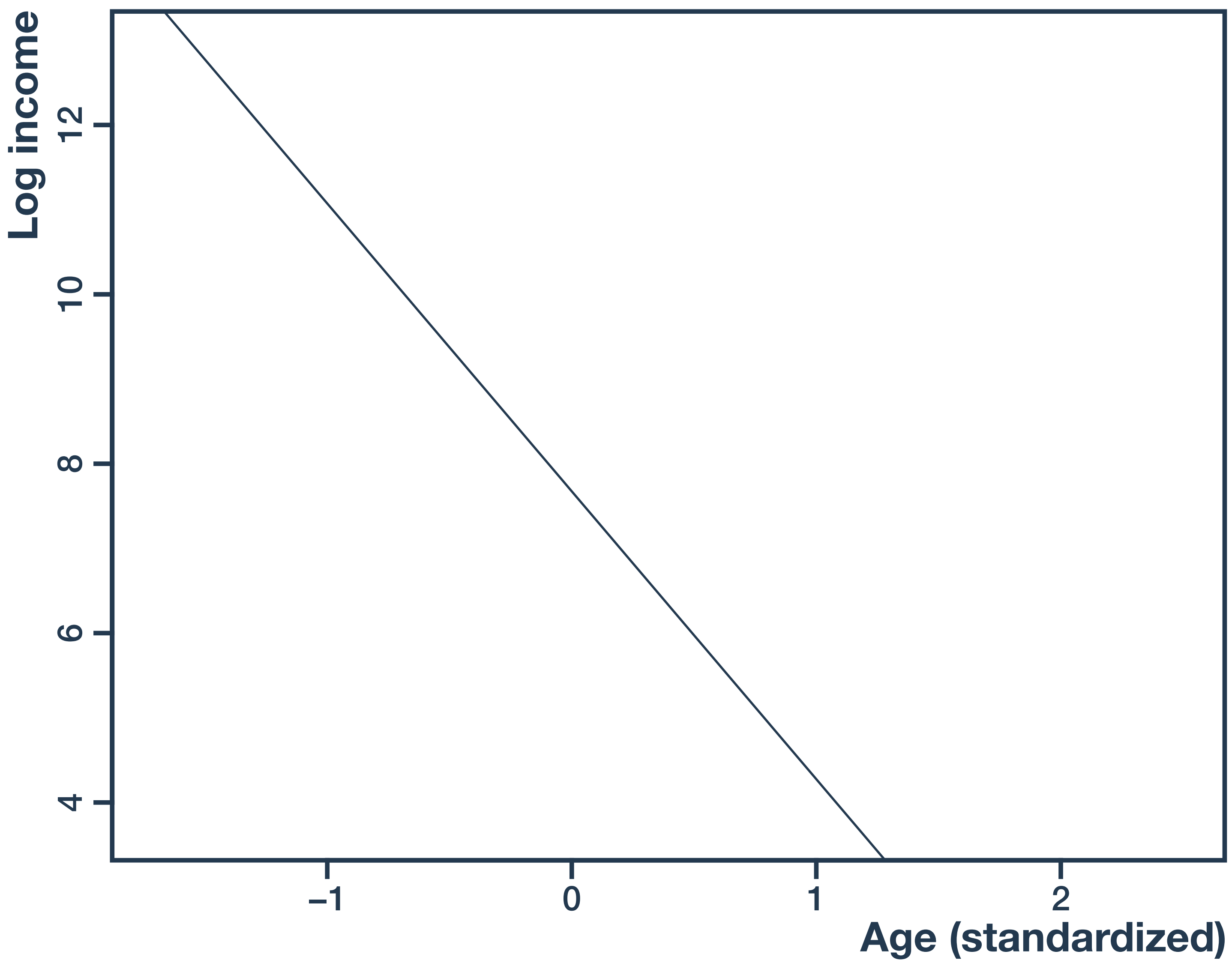
$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$

$$\beta \sim \text{Norm}(0, 3)$$

$$\sigma \sim \text{Unif}(0, 5)$$



α	β
7.674	-3.399

Prior predictive plot

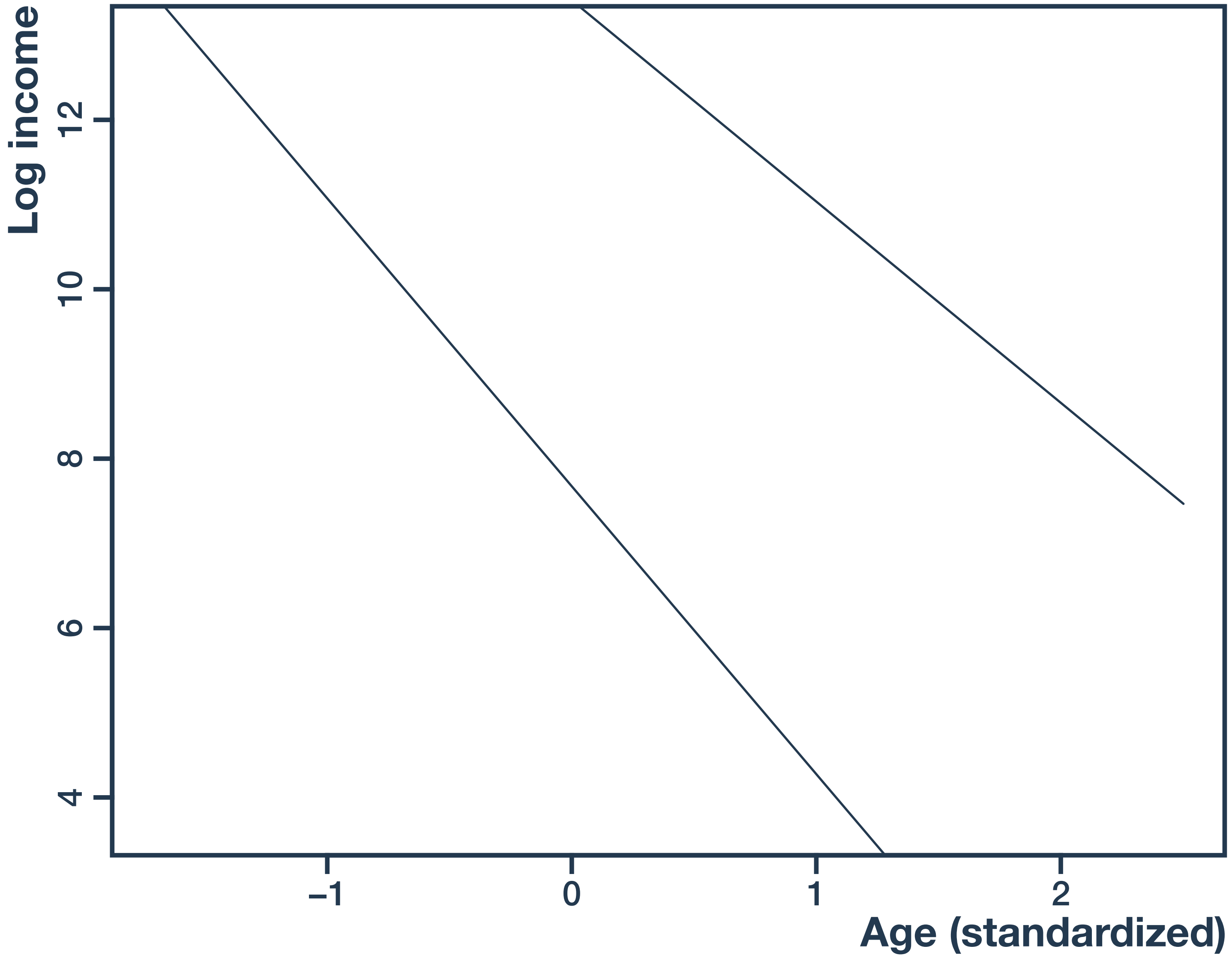
$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$

$$\beta \sim \text{Norm}(0, 3)$$

$$\sigma \sim \text{Unif}(0, 5)$$

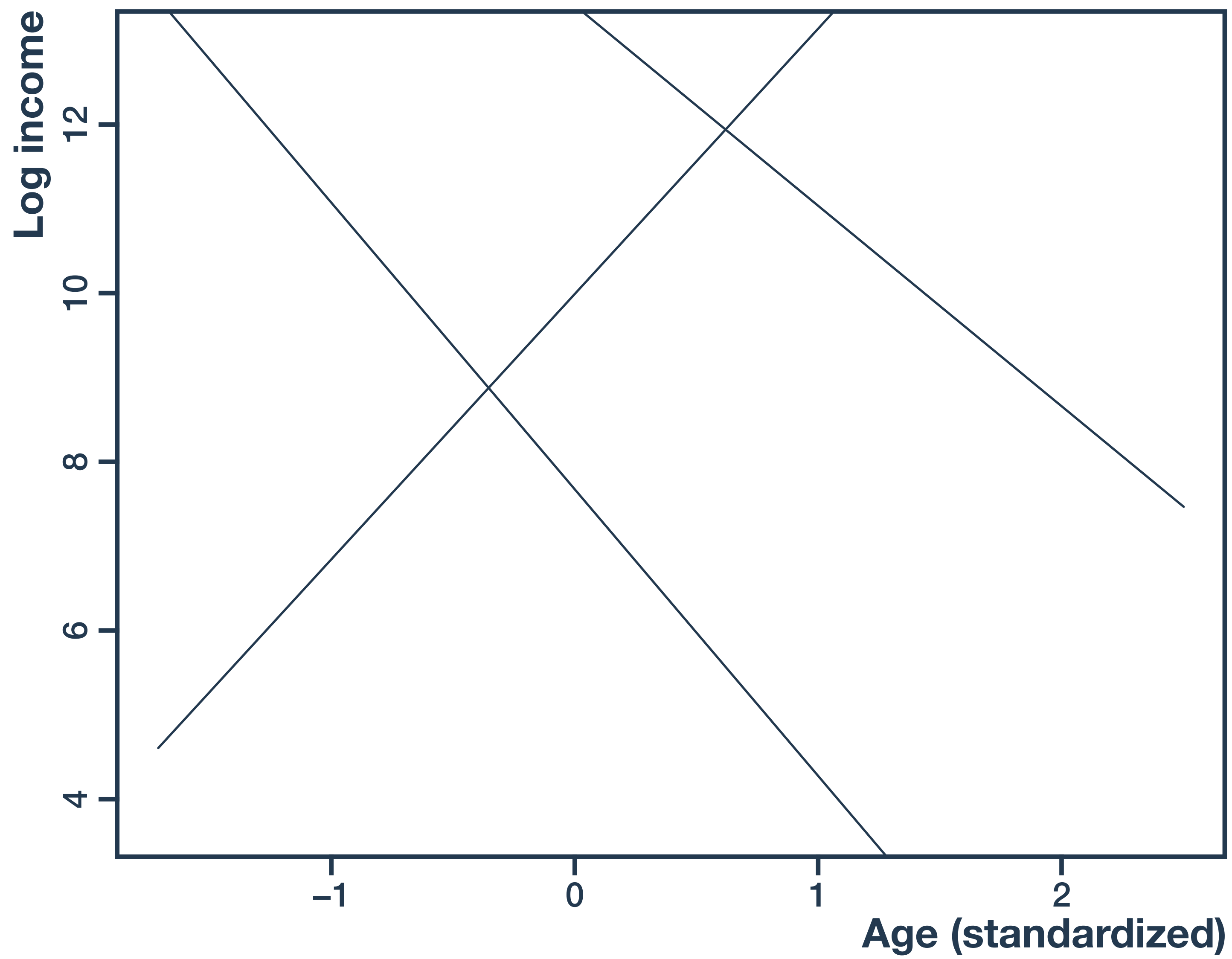


α	β
7.674	-3.399
13.412	-2.376

Prior predictive plot

$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$
$$\beta \sim \text{Norm}(0, 3)$$
$$\sigma \sim \text{Unif}(0, 5)$$



α	β
7.674	-3.399
13.412	-2.376
9.990	3.146

Prior predictive plot

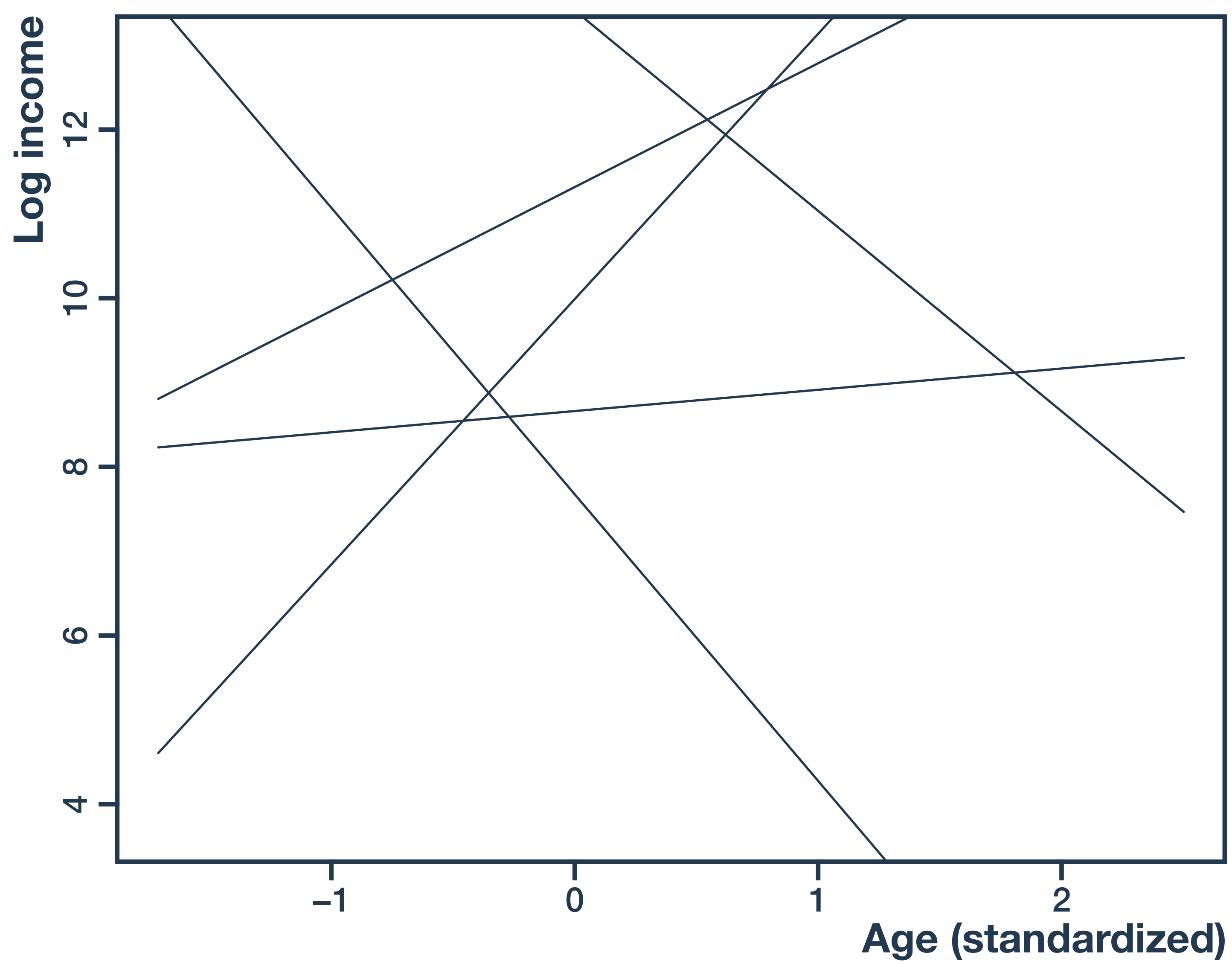
$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$

$$\beta \sim \text{Norm}(0, 3)$$

$$\sigma \sim \text{Unif}(0, 5)$$

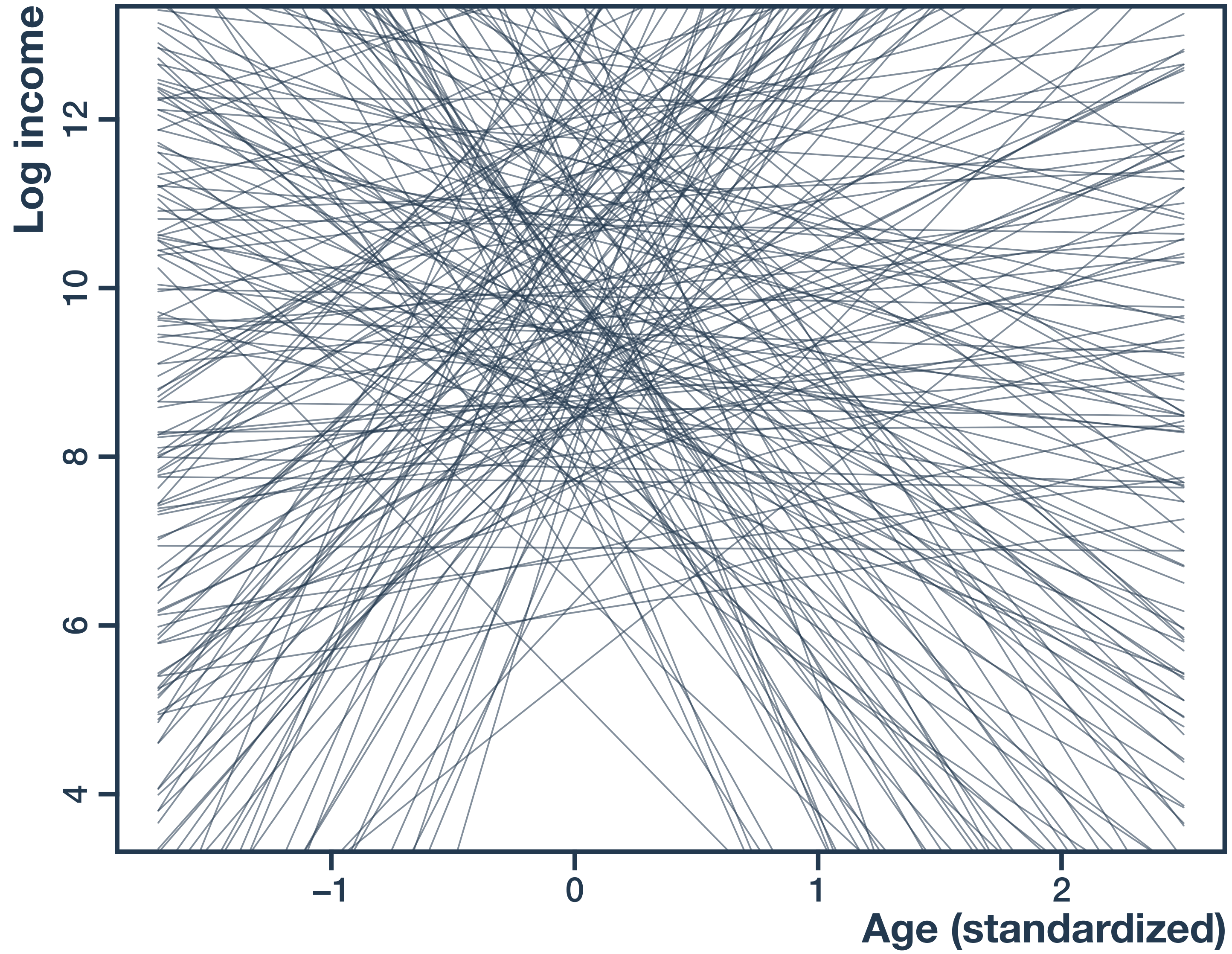


α	β
7.674	-3.399
13.412	-2.376
9.990	3.146
11.318	1.468
8.662	0.252

Prior predictive plot

$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$
$$\beta \sim \text{Norm}(0, 3)$$
$$\sigma \sim \text{Unif}(0, 5)$$

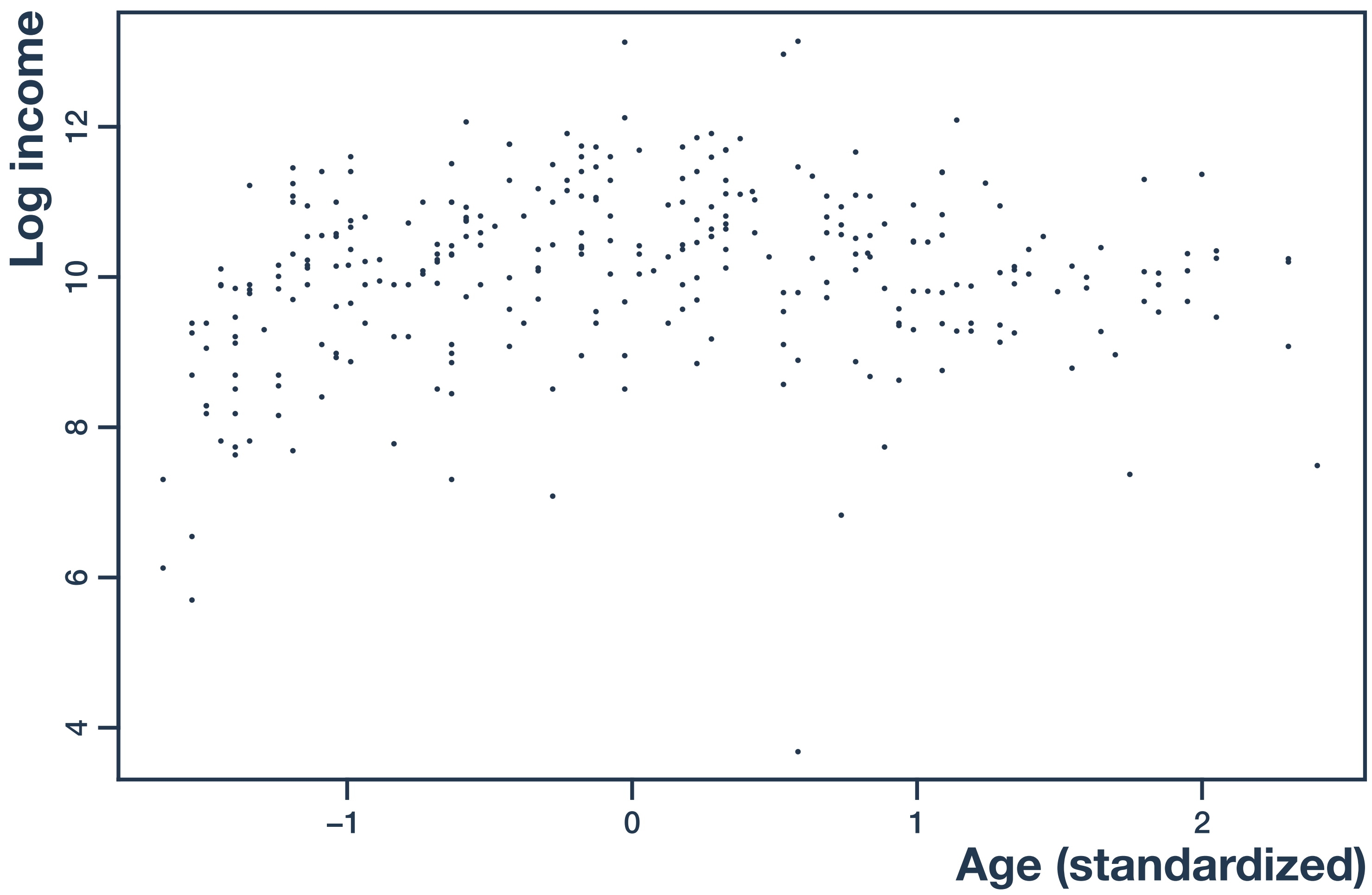


α	β
7.674	-3.399
13.412	-2.376
9.990	3.146
11.318	1.468
8.662	0.252
⋮	⋮

Visualizing predictions

$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$
$$\beta \sim \text{Norm}(0, 3)$$
$$\sigma \sim \text{Unif}(0, 5)$$

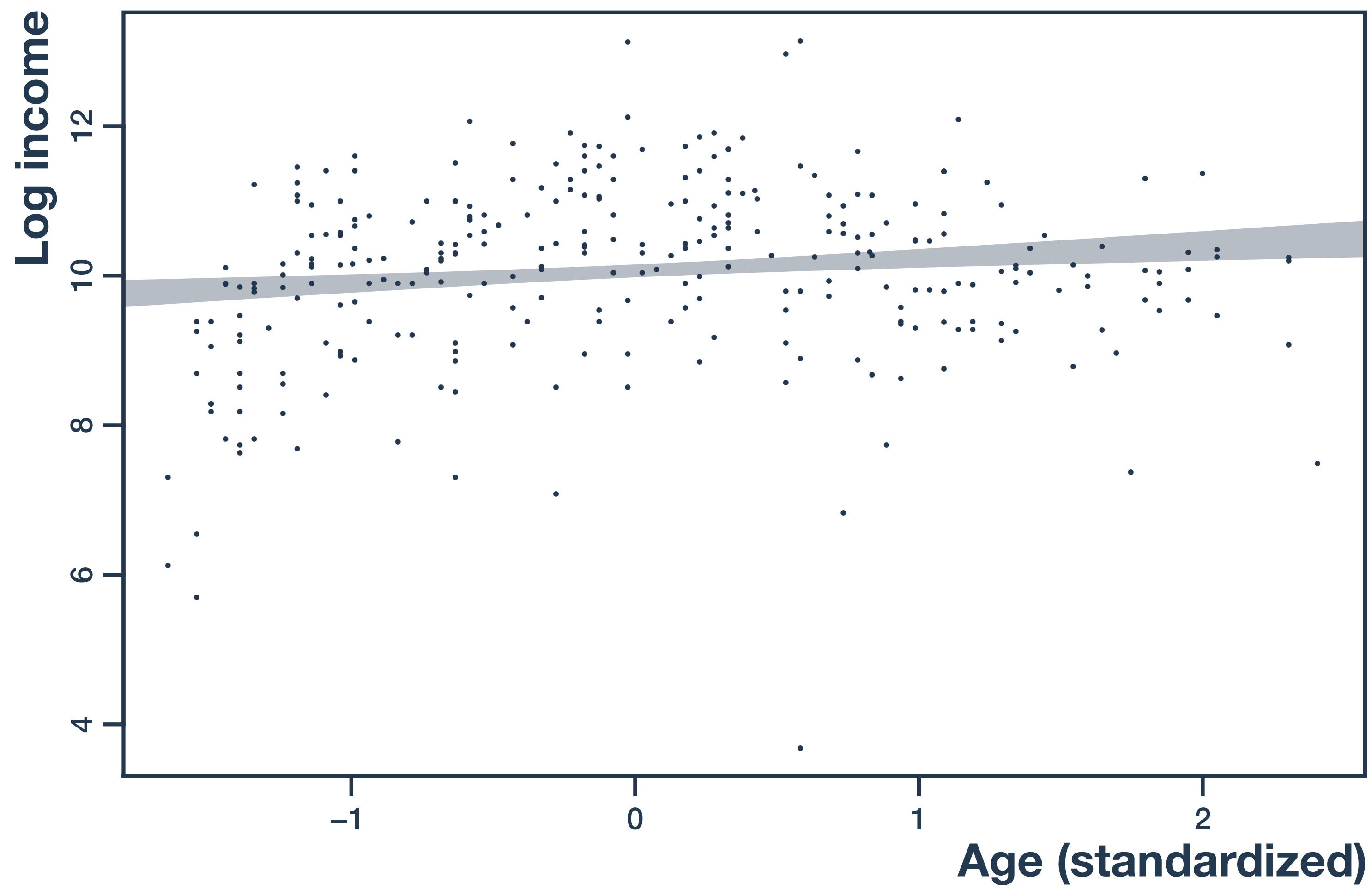


Visualizing predictions

$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$
$$\beta \sim \text{Norm}(0, 3)$$
$$\sigma \sim \text{Unif}(0, 5)$$

	Post. Mean	exp(Mean)
α	10.06	0.07
β	0.17	0.07
σ	1.18	0.05



Posterior distribution of mean:

$$\Pr(\mu | \text{Age} = a)$$

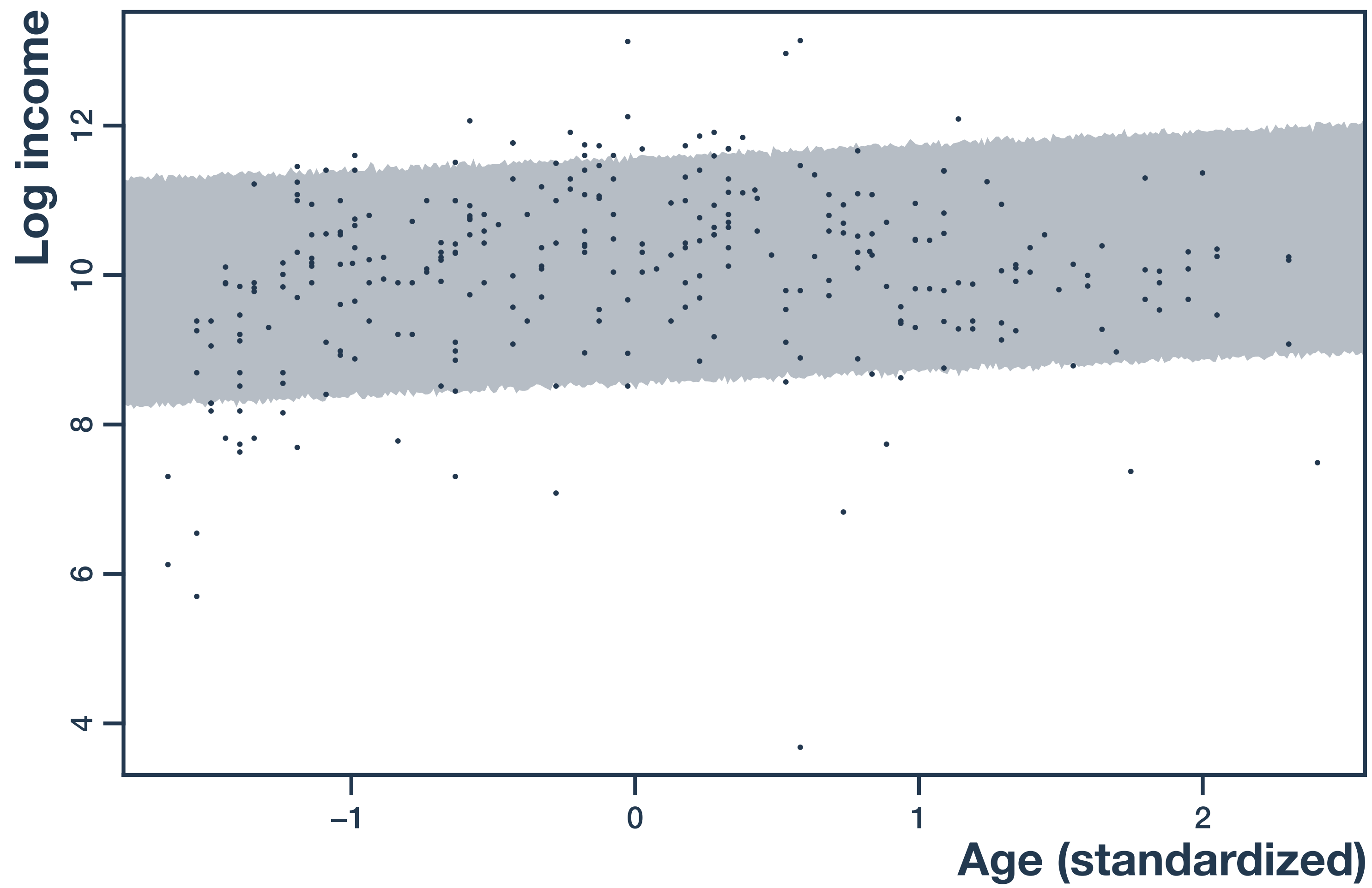
1. Take a sample of size N from posterior Pr(α, β, σ|D).
2. For each value of Age a, calculate N values of $\mu = \alpha + \beta a$.
3. Calculate quantiles (say, 10% and 90%) for posterior of μ at each value of a.

Visualizing predictions

$$\log(\text{Inc}_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta \text{St}(\text{Age}_i)$$

$$\alpha \sim \text{Norm}(10, 2)$$
$$\beta \sim \text{Norm}(0, 3)$$
$$\sigma \sim \text{Unif}(0, 5)$$

	Post. Mean	exp(Mean)
α	10.06	0.07
β	0.17	0.07
σ	1.18	0.05



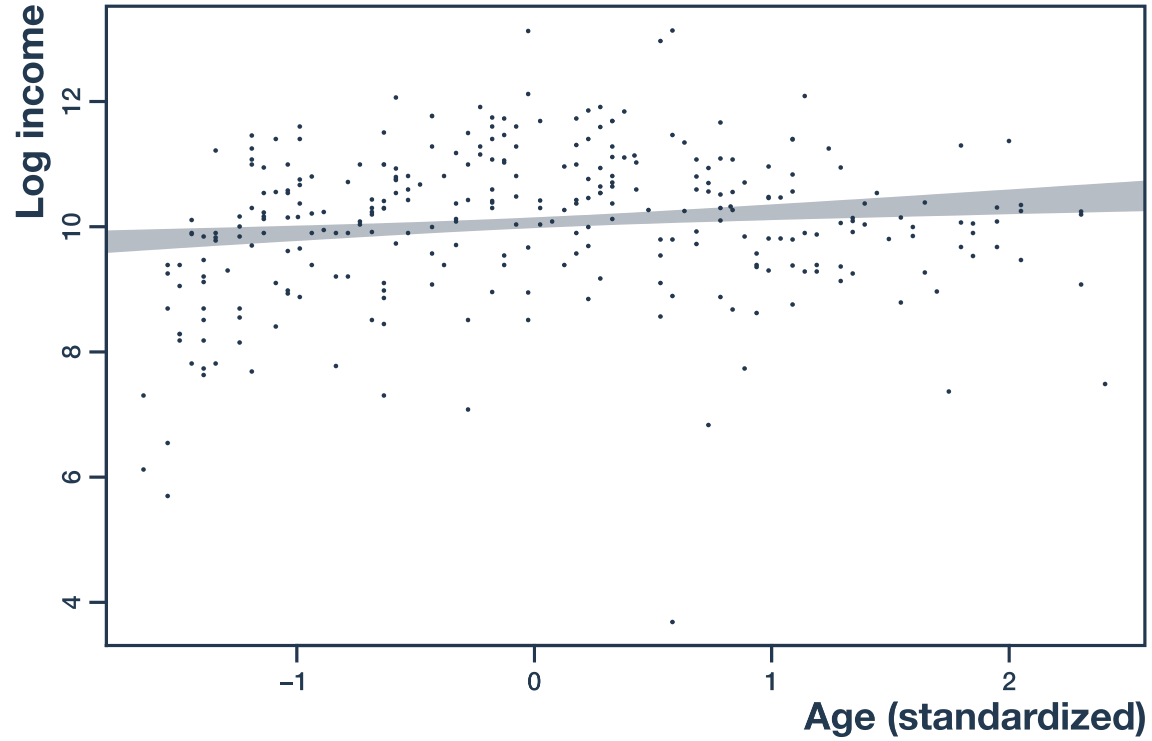
Posterior predictive distribution:

$$\Pr(\log(Inc) | \text{Age} = a)$$

1. Take a sample of size N from posterior Pr(α, β, σ|D).
2. For each value of Age a, calculate N values of $\mu = \alpha + \beta a$.
3. Draw from Norm(μ, σ) for each of the N posterior samples.
4. Calculate quantiles of these predicted outcomes.

Mean versus prediction

Posterior distribution of mean:
Pr(μ | Age, D)

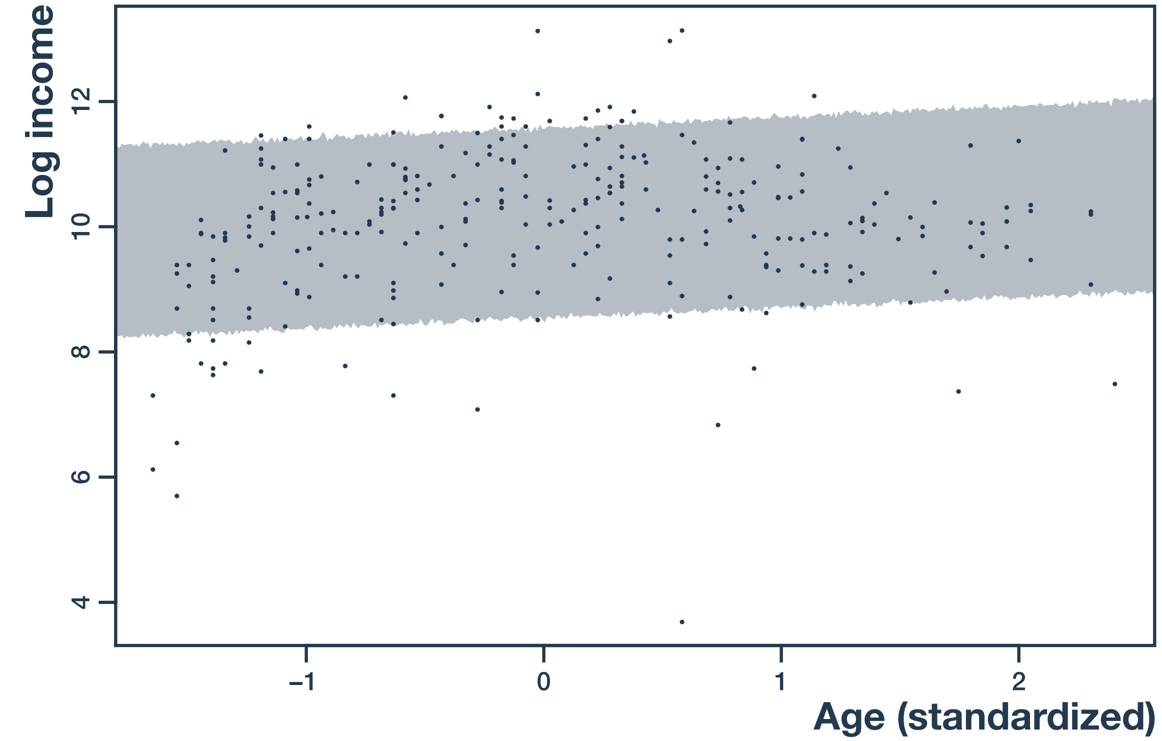


For any given age, μ is the "expected" (mean) log income for people of that age.

The posterior distribution of μ describes our modeled uncertainty about the value of μ.

This distribution takes into account coefficients α and β, but not the standard deviation σ.

Posterior *predictive* distribution:
Pr(log(Inc) | Age, D)



For any given age, the posterior predictive distribution predicts the log income for any individual person of that age.

The posterior distribution of μ describes our modeled uncertainty about the value of log(Inc).

This distribution takes into account coefficients α, β, and σ.

The 80% posterior interval should contain about 80% of the data.

Assessing fit