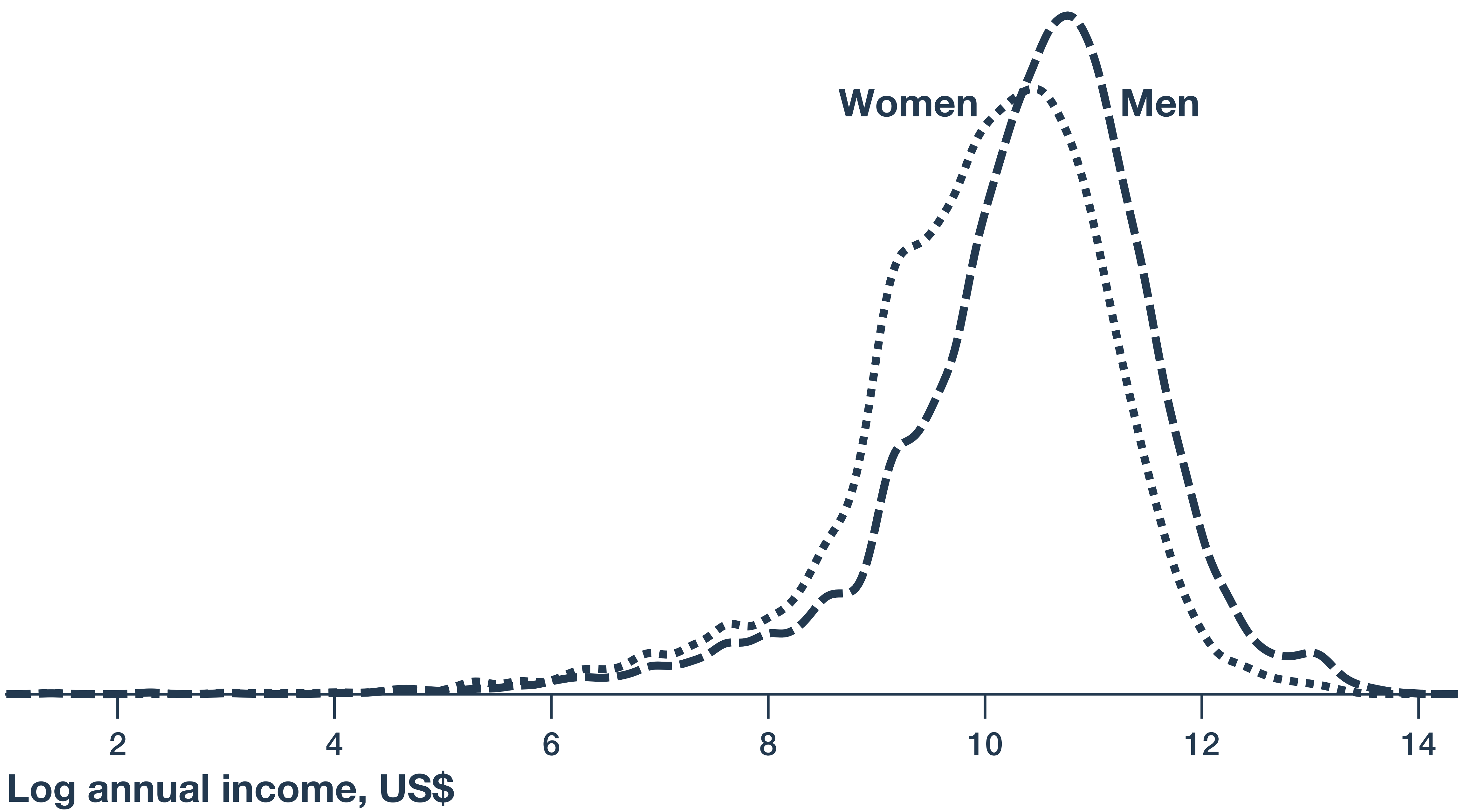
Agenda

1. **Linear regression with one covariate**
2. **Joint posteriors**
3. **Interpreting coefficients at log-scale**
4. **Linear regression with many covariates**
5. **Estimation and working with samples in R**

Modeling income by gender



Modeling income by gender

$$y_i \sim \mathrm{Norm}(\mu, \sigma)$$

Modeling income by gender

Value of μ depends on the person

$$y_i \sim \mathrm{Norm}(\mu_i, \sigma)$$

Modeling income by gender

$$y_i \sim \mathrm{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta w_i$$

$\mu_i = \alpha$ for men

$\mu_i = \alpha + \beta$ for women

Modeling income by gender

$$y_i \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta w_i$$

$$\alpha \sim \text{Norm}(0, 30)$$

Prior for each parameter

$$\beta \sim \text{Norm}(0, 30)$$

$$\sigma \sim \text{Unif}(0, 50)$$

Modeling income by gender

No predictors

$$y_i \sim \text{Norm}(\mu, \sigma)$$
$$\mu \sim \text{Norm}(0, 30)$$
$$\sigma \sim \text{Unif}(0, 50)$$

One predictor

$$y_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta w_i$$
$$\alpha \sim \text{Norm}(0, 30)$$
$$\beta \sim \text{Norm}(0, 30)$$
$$\sigma \sim \text{Unif}(0, 50)$$

Alternate expressions

One model,
three representations

$y_i \sim \text{Norm}(\mu_i, \sigma)$		$y_i = \alpha + \beta w_i + \varepsilon_i$
$\mu_i = \alpha + \beta w_i$	$y_i \sim \text{Norm}(\alpha + \beta w_i, \sigma)$	$\varepsilon_i \sim \text{Norm}(0, \sigma)$
$\alpha \sim \text{Norm}(0, 30)$	$\alpha \sim \text{Norm}(0, 30)$	$\alpha \sim \text{Norm}(0, 30)$
$\beta \sim \text{Norm}(0, 30)$	$\beta \sim \text{Norm}(0, 30)$	$\beta \sim \text{Norm}(0, 30)$
$\sigma \sim \text{Unif}(0, 50)$	$\sigma \sim \text{Unif}(0, 50)$	$\sigma \sim \text{Unif}(0, 50)$

Joint posterior

When we estimate this model, we get a single **joint** posterior distribution for **all three** parameters:

$$\Pr\left(\alpha, \beta, \sigma \middle| D\right)$$

What can we do with a joint posterior?

Working with the posterior

$$\Pr(\alpha, \beta, \sigma | D)$$

Data:
Sample of 35,124 working adults in the United States

1. Describe the marginal posterior distributions

$$\Pr(\alpha | D), \Pr(\beta | D), \Pr(\sigma | D)$$

	Mean	Std. Dev.	2.5%	97.5%
α	10.382	0.009	10.364	10.400
β	–0.434	0.013	-0.459	-0.408
σ	1.221	0.005	1.212	1.230

Working with the posterior

$$\Pr(\alpha, \beta, \sigma | D)$$

Data:
Sample of 35,124 working adults in the United States

1. Describe the marginal posterior distributions

   $\Pr(\alpha | D), \Pr(\beta | D), \Pr(\sigma | D)$

2. Describe posterior probability of theoretically relevant scenarios

   $\Pr(\beta < 0 | D)$

$$y_i \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta w_i$$
$$\alpha \sim \text{Norm}(0, 30)$$
$$\beta \sim \text{Norm}(0, 30)$$
$$\sigma \sim \text{Unif}(0, 50)$$

Working with the posterior

$$\Pr(\alpha, \beta, \sigma | D)$$

Data:
Sample of 35,124 working adults in the United States

1. Describe the marginal posterior distributions

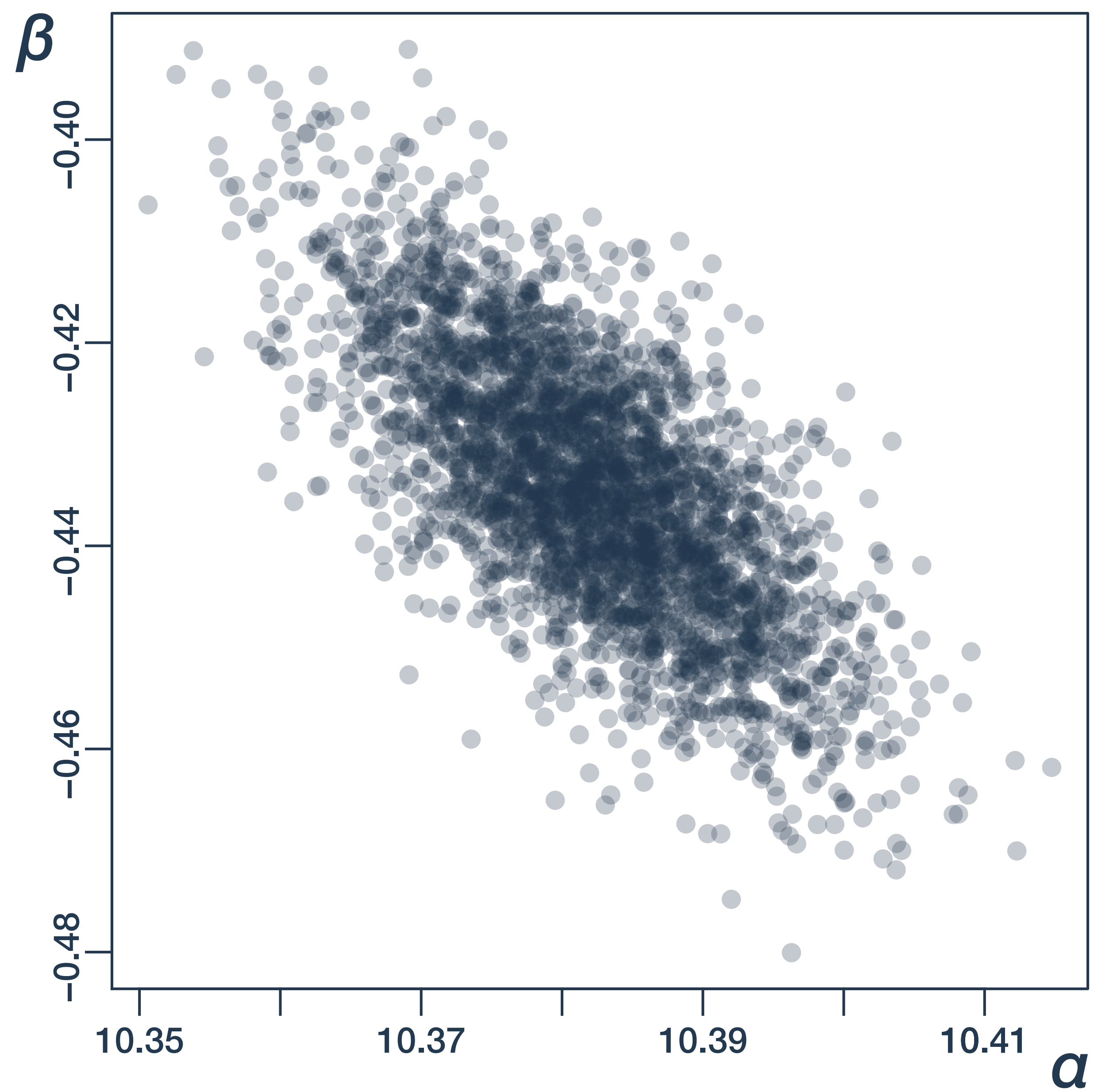
   $\Pr(\alpha|D), \Pr(\beta|D), \Pr(\sigma|D)$

2. Describe posterior probability of theoretically relevant scenarios

   $\Pr(\beta < 0|D)$

3. **Describe the partial joint posterior distribution**

   $\Pr(\alpha, \beta|D)$



Interpreting log-scale coefficients

y_i = log(income) $\longrightarrow$ $$y_i \sim \mathrm{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta w_i$$ $\longleftarrow$

$\mu_i = \alpha$ for men

$\mu_i = \alpha + \beta$ for women

	Mean	Std. Dev.	2.5%	97.5%
α	10.382	0.009	10.364	10.400
β	–0.434	0.013	-0.459	-0.408
σ	1.221	0.005	1.212	1.230

$\mu_m = 10.382 \approx \log(32{,}300)$

$\mu_w = 9.948 \approx \log(20{,}900)$

In general: if the outcome variable is on a log-scale, then exponentiating coefficient estimates (e^{α}) gives *multiplicative* factors

$\exp(-0.434) \approx 0.65$:
These results suggest that women make about 35% less than men on average

Modeling income

Adding covariates

$$
\begin{aligned}
y_i &\sim \text{Norm}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_1 w_i + \beta_2 age_i + \beta_3 college_i \\
\alpha &\sim \text{Norm}(0, 30) \\
\beta_1 &\sim \text{Norm}(0, 30) \\
\beta_2 &\sim \text{Norm}(0, 30) \\
\beta_3 &\sim \text{Norm}(0, 30) \\
\sigma &\sim \text{Unif}(0, 50)
\end{aligned}
$$

---

compact notation:

$$
\begin{aligned}
y_i &\sim \text{Norm}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_1 w_i + \beta_2 age_i + \beta_3 college_i \\
\alpha, \beta_1, \beta_2, \beta_3 &\sim \text{Norm}(0, 30) \\
\sigma &\sim \text{Unif}(0, 50)
\end{aligned}
$$