R2D2 prior

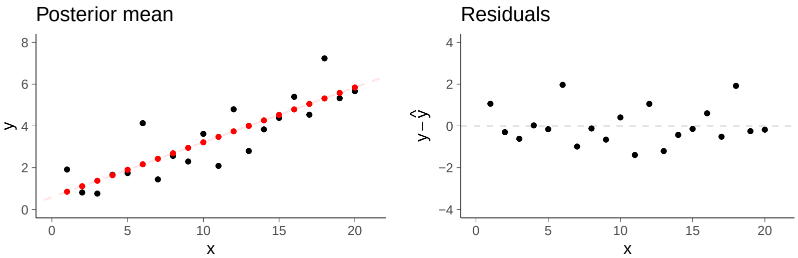
- Useful for models with many predictors

R^2, proportion of variance explained

$$\frac{\text{explained variance}}{\text{total variance}}$$

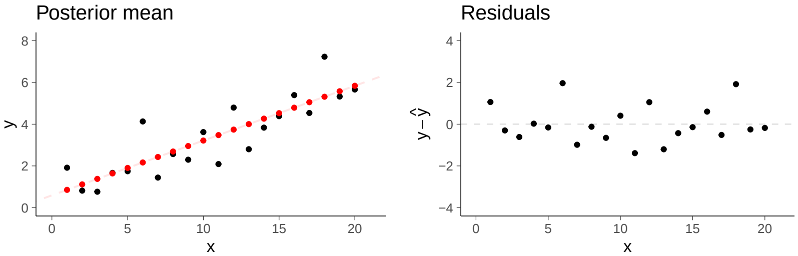
R^2, proportion of variance explained

$$\frac{\text{explained variance}}{\text{total variance}}$$



R^2, proportion of variance explained

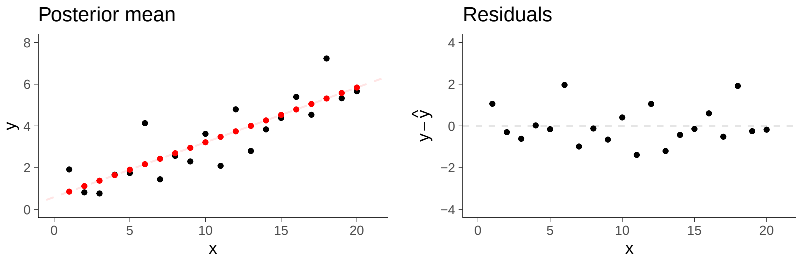
$$\frac{\text{explained variance}}{\text{total variance}}$$



$$R^2 = \frac{\mathrm{Var}_\mu}{\mathrm{Var}_\mu + \mathrm{Var}_{res}}, \text{ where } \mu = \hat{y}, \text{and res} = y - \hat{y}$$

R^2, proportion of variance explained

$$\frac{\text{explained variance}}{\text{total variance}}$$



$$R^2 = \frac{\mathrm{Var}_\mu}{\mathrm{Var}_\mu + \mathrm{Var}_{res}}, \text{ where } \mu = \hat{y}, \text{ and res} = y - \hat{y}$$

$$\approx \frac{2.4}{2.4 + 0.85} \approx 0.74$$

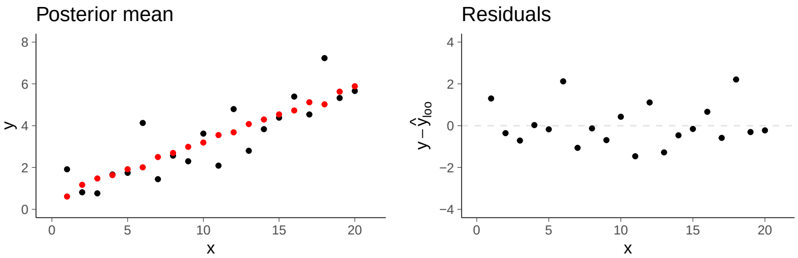
LOO-R^2

Replace posterior mean prediction with LOO-CV predictions

Code and examples avehtari.github.io/bayes_R2/bayes_R2.html

LOO-R^2

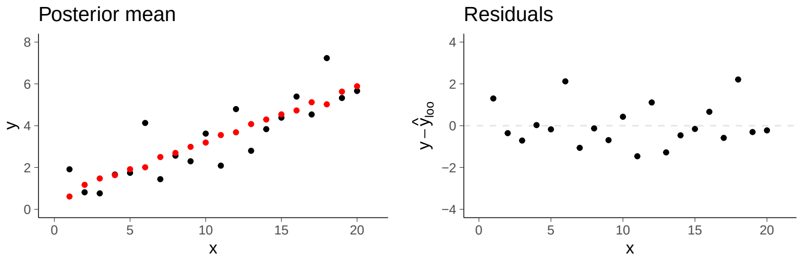
Replace posterior mean prediction with LOO-CV predictions



Code and examples avehtari.github.io/bayes_R2/bayes_R2.html

LOO-R^2

Replace posterior mean prediction with LOO-CV predictions



$$R^2 \approx \frac{2.4}{2.4 + 0.85} \approx 0.74$$

$$\text{LOO}-R^2 \approx \frac{2.4}{2.4 + 1.0} \approx 0.7$$

Code and examples avehtari.github.io/bayes_R2/bayes_R2.html

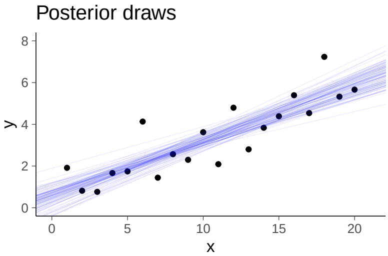
Bayesian-R^2

Model based, with posterior

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Bayesian-R^2

Model based, with posterior

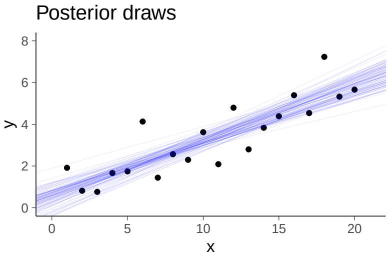


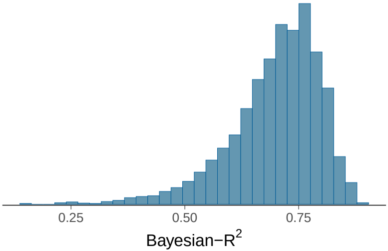
$$\mathrm{Var}_{\mu}^{s} = V_{n=1}^{N}\hat{y}_n^s$$
$$\mathrm{Var}_{\mathrm{res}}^{s} = \mathrm{Var}(y \mid x, \alpha^s, \beta^s, \sigma^s) = (\sigma^2)^s$$

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Bayesian-R^2

Model based, with posterior





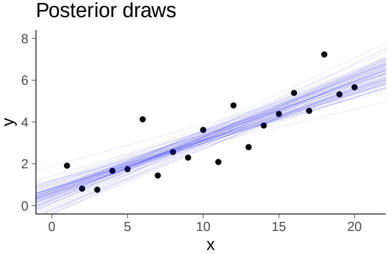
$$\mathrm{Var}_{\mu}^{s} = V_{n=1}^{N} \hat{y}_n^s$$

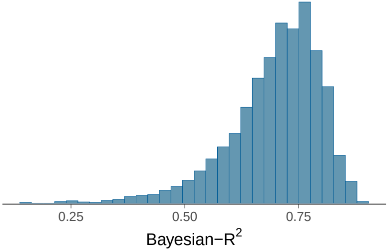
$$\mathrm{Var}_{\mathrm{res}}^{s} = \mathrm{Var}(y \mid x, \alpha^s, \beta^s, \sigma^s) = (\sigma^2)^s$$

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Bayesian-R^2

Model based, with posterior





$$\mathrm{Var}^s_{\mu} = V^N_{n=1} \hat{y}^s_n$$

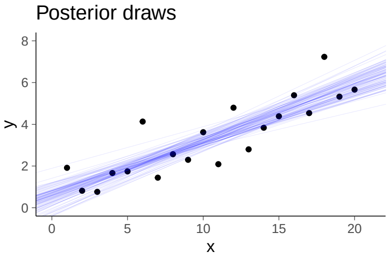
$$\mathrm{Var}^s_{\mathrm{res}} = \mathrm{Var}(y \mid x, \alpha^s, \beta^s, \sigma^s) = (\sigma^2)^s$$

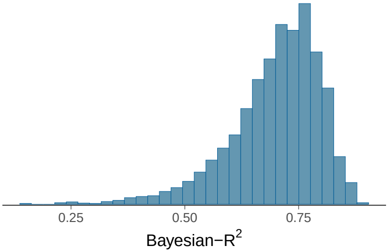
- Posterior presents the uncertainty

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Bayesian-R^2

Model based, with posterior





$$\mathrm{Var}_{\mu}^{s} = V_{n=1}^{N} \hat{y}_n^s$$

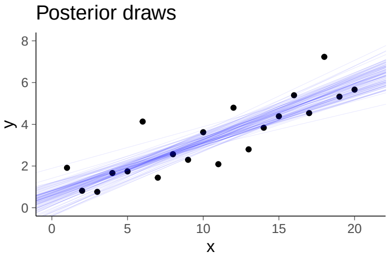
$$\mathrm{Var}_{res}^{s} = \mathrm{Var}(y \mid x, \alpha^s, \beta^s, \sigma^s) = (\sigma^2)^s$$

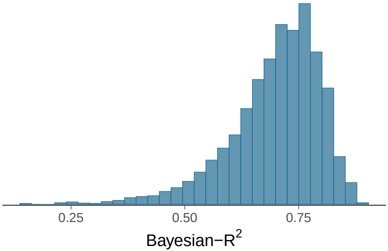
- Posterior presents the uncertainty
- Posterior mean of Bayesian-R^2 is optimistic compared to LOO-R^2

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Bayesian-R^2

Model based, with posterior





$$\mathrm{Var}_{\mu}^{s} = V_{n=1}^{N}\hat{y}_{n}^{s}$$

$$\mathrm{Var}_{res}^{s} = \mathrm{Var}(y \mid x, \alpha^{s}, \beta^{s}, \sigma^{s}) = (\sigma^{2})^{s}$$

- Posterior presents the uncertainty
- Posterior mean of Bayesian-R^2 is optimistic compared to LOO-R^2
- Can also be used to examine R^2 prior distribution

Gelman, Goodrich, Gabry, and Vehtari (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307-309.

Implied prior on R^2

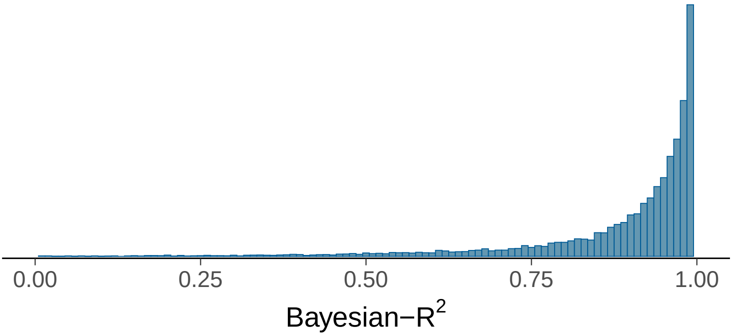
- Draw from the model parameter priors
- Compute prior predictions
- Compute Bayesian-R^2 using the prior draws

Implied prior on R^2

Regression and Other Stories, Section 12.7 Models for regression coefficients: 26 covariates normalized to have mean 0 and sd 1

Independent normal priors for coefficients

$$\beta \sim \text{normal}(0, 2.5)$$
$$\sigma \sim t_3(0, 3)$$

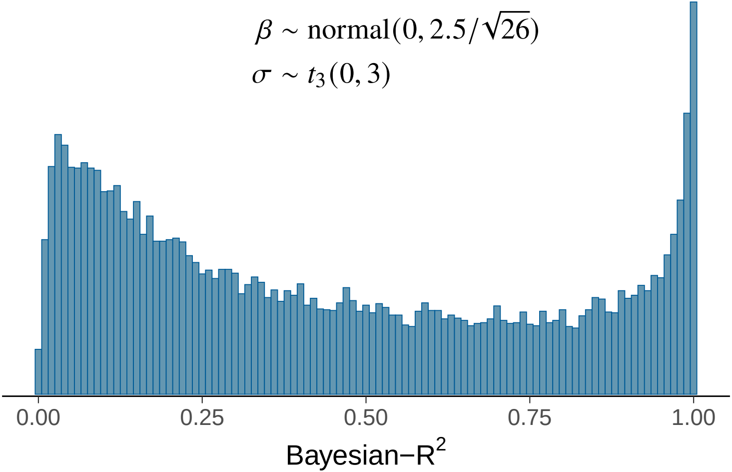


Implied prior on R^2

Regression and Other Stories, Section 12.7 Models for regression coefficients: 26 covariates normalized to have mean 0 and sd 1

Scaled normal priors for coefficients

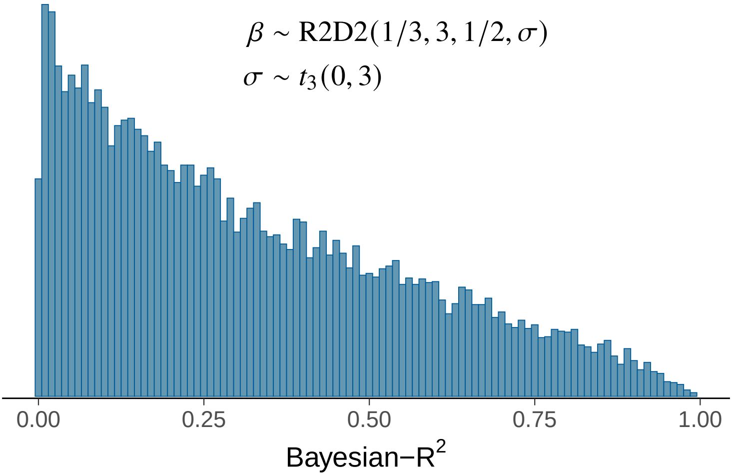
$$\beta \sim \text{normal}(0, 2.5/\sqrt{26})$$
$$\sigma \sim t_3(0, 3)$$



Implied prior on R^2

Regression and Other Stories, Section 12.7 Models for regression coefficients: 26 covariates normalized to have mean 0 and sd 1

R2D2 prior



R2D2 prior

- Define the prior on R^2
- Joint prior on coefficients which depends also on σ
- D2 refers to Dirichlet decomposition

Zhang, Naughton, Bondell, and Reich (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117:862–874.

Aguilar and Bürkner (2023). Intuitive joint priors for Bayesian linear multilevel models: The R2D2M2 prior. *Electronic Journal of Statistics*, 17:1711–1767.

R2D2 prior

- Define the prior on R^2
- Joint prior on coefficients which depends also on σ
- D2 refers to Dirichlet decomposition
- With mean 1/3 and precision 3, R2D2 prior states smaller R^2 are more likely a priori, but locally the slope is not steep

Zhang, Naughton, Bondell, and Reich (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117:862–874.

Aguilar and Bürkner (2023). Intuitive joint priors for Bayesian linear multilevel models: The R2D2M2 prior. *Electronic Journal of Statistics*, 17:1711–1767.

R2D2 prior

- Define the prior on R^2
- Joint prior on coefficients which depends also on σ
- D2 refers to Dirichlet decomposition
- With mean 1/3 and precision 3, R2D2 prior states smaller R^2 are more likely a priori, but locally the slope is not steep
- Often good prior when there are many covariates

Zhang, Naughton, Bondell, and Reich (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117:862–874.

Aguilar and Bürkner (2023). Intuitive joint priors for Bayesian linear multilevel models: The R2D2M2 prior. *Electronic Journal of Statistics*, 17:1711–1767.

R2D2 prior

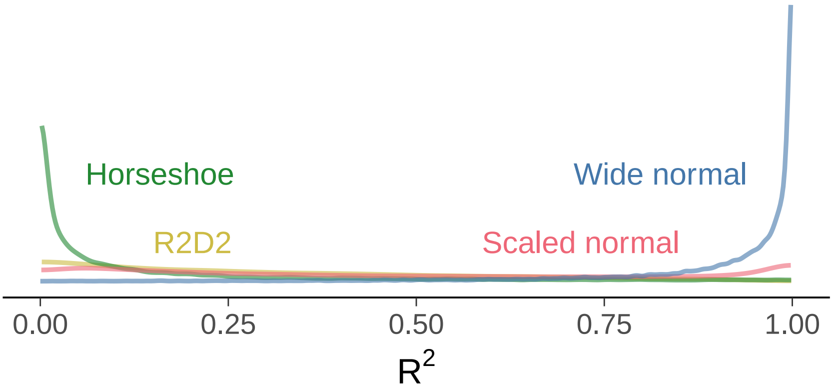
- Define the prior on R^2
- Joint prior on coefficients which depends also on σ
- D2 refers to Dirichlet decomposition
- With mean 1/3 and precision 3, R2D2 prior states smaller R^2 are more likely a priori, but locally the slope is not steep
- Often good prior when there are many covariates
- Predictively consistent prior: adding more covariates doesn't change the prior on R^2

Zhang, Naughton, Bondell, and Reich (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117:862–874.

Aguilar and Bürkner (2023). Intuitive joint priors for Bayesian linear multilevel models: The R2D2M2 prior. *Electronic Journal of Statistics*, 17:1711–1767.

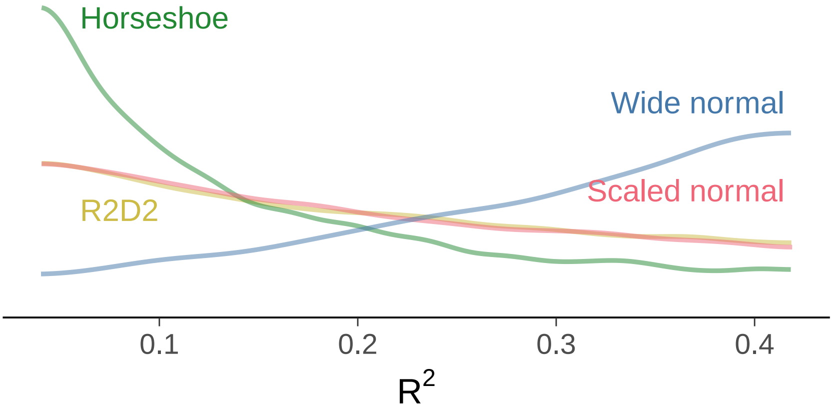
Priors and posteriors

a) Implied prior on R^2

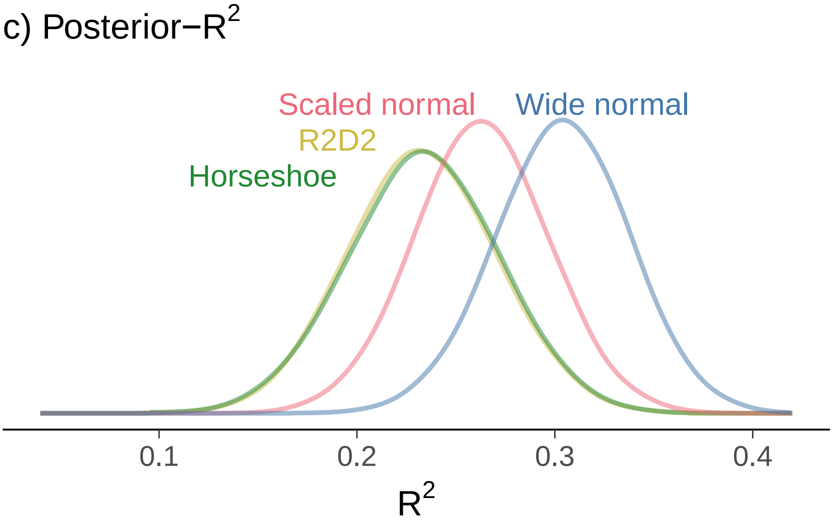


Priors and posteriors

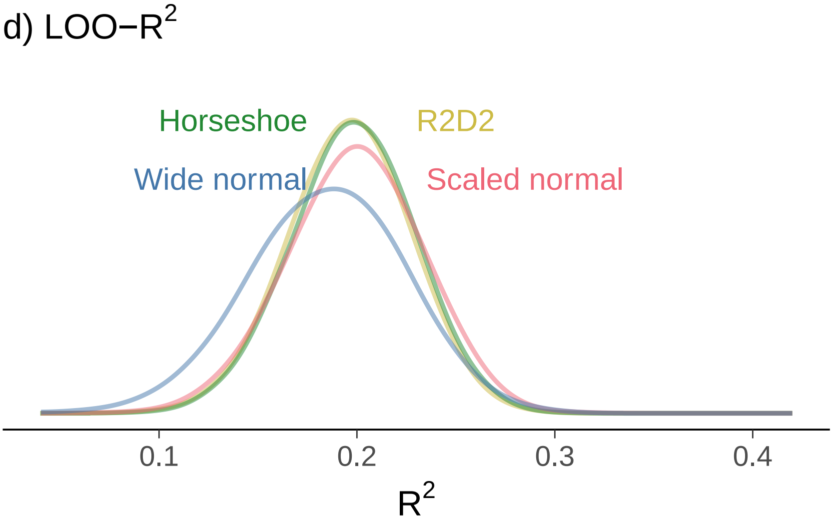
b) Implied prior on R^2



Priors and posteriors



Priors and posteriors



brms

```
fit <- brm(y ~ .,
           prior = prior(R2D2(mean_R2 = 1/3,
                              prec_R2 = 3,
                              cons_D2 = 1/2),
                         class=b)))
```

brms

```
fit <- brm(y ~ .,
           prior = prior(R2D2(mean_R2 = 1/3,
                              prec_R2 = 3,
                              cons_D2 = 1/2),
                         class=b)))
```

See the assignment notebook for additional R^2 related code.