Bayesian multiple regression

Dr. Merlise Clyde

Duke
UNIVERSITY

cognitive scores

model $kid\ score_i = \beta_{intercept} + \beta_{hs} hs_i + \beta_{iq} iq_i + \beta_{work} work_i + \beta_{age} age_i + \varepsilon_i$

$$\varepsilon_i \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma^2)$$

- prior distribution for $\beta_{intercept}$, β_{hs}, β_{iq}, β_{work}, β_{age} and σ^2
- conjugate family multivariate normal-gamma
- mean, variance and covariance of all regression coefficients
- prior parameters of the gamma distribution of $1/\sigma^2$

reference prior

Data reference: http://www.stat.columbia.edu/~gelman/arm/examples/child.iq/

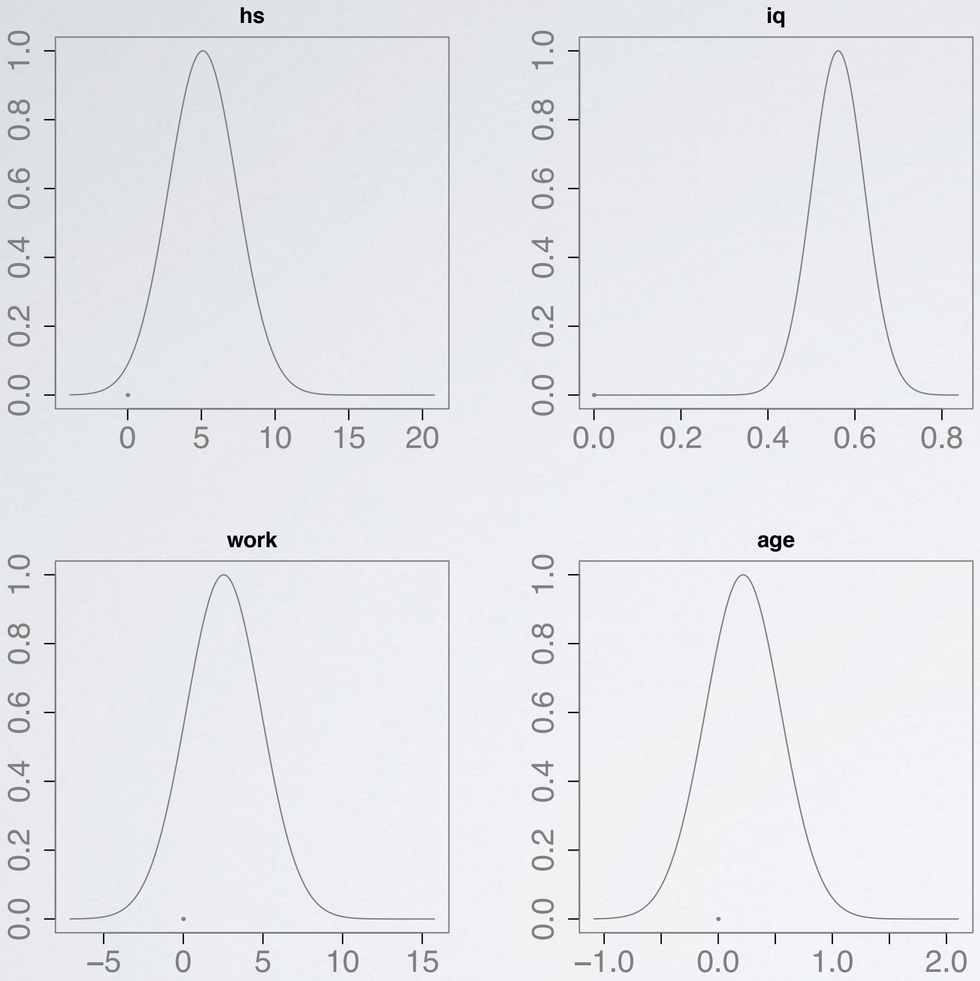
reference analysis

- prior distributions: uniform for β_j and $p(\sigma^2) \propto 1/\sigma^2$
- posterior distribution

$$\beta_j \mid y_1, \ldots, y_n \sim t_{n-p-1}\left(\hat{\beta}_j, \mathsf{SE}(\beta_j)^2\right)$$

- degrees of freedom $n - p - 1$
- posterior mean $\hat{\beta}_j$
- posterior standard deviation $\mathsf{SE}(\beta_j)$

example


hs
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20

iq
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8

work
0.0
0.2
0.4
0.6
0.8
1.0
−5
0
5
10
15

age
0.0
0.2
0.4
0.6
0.8
1.0
−1.0
0.0
1.0
2.0

highest posterior density intervals

	post. mean	post. sd	2.5%	97.5%
hs	5.09	2.31	0.55	9.64
iq	0.56	0.06	0.44	0.68
work	2.54	2.35	-2.08	7.16
age	0.22	0.33	-0.43	0.87

posterior mean $\beta_j \pm t_{1-\alpha/2,n-p-1}$ sd(β_j)

summary

- reference analysis
- use standard software to obtain
- interpretation of credible intervals

next video

- model selection