EXERCISE 9.21. *FEV Data Analysis Project*. Conduct a complete Bayesian linear regression analysis of the *full* FEV data available from our website that includes as candidate predictor variables Age, Smoke, Height (in inches), and Sex (1 = male). One issue to be addressed in this study is the determination of normal ranges of FEV for a given type of adolescent. For instance, what FEV is predicted for a 15-year-old male who does not smoke and is 66 inches tall? Do the following in your analysis.

(a) An exploratory data analysis.

(b) Discuss prior construction, predictor selection (consider interactions and higher order terms), and convergence and model diagnostics.

(c) Present posterior inferences for regression parameters and for subpopulation means in appropriately designed tables or figures. Based on your analysis, is smoking related to FEV?

(d) Determine normal FEV ranges for several different types of adolescents presenting the results in a table.

(e) Discuss a sensitivity analysis.

(f) Write-up of your entire analysis.

Dr. David Mannino, M.D. (Division of Pulmonary, Critical Care, and Sleep Medicine and Director of the Pulmonary Epidemiology Research Laboratory at the University of Kentucky) provided prior information. The values are measured in liters. The two numbers are the prior best guess of the mean FEV followed by the 99th percentile for the mean. For 18-year-old, female smokers, 70 inches tall: 4.0 and 4.8. For 16-year-old, male nonsmokers, 70 inches tall: 4.2 and 5.0. For 13-year-old, male smokers, 66 inches tall: 3.4 and 4.0. For 12-year-old, male nonsmokers, 60 inches tall: 2.7 and 3.5.

EXERCISE 9.22. *The Coleman Report Data*. Mosteller and Tukey (1977) and Christensen (1996) reproduced data collected from schools in the New England and Mid-Atlantic states of the USA. Consider two variables: y – the mean verbal test score for sixth graders and x – a composite measure of socioeconomic status associated with the school. The data are presented in Table 9.7. We wish to predict y based on x. Conduct a complete Bayesian regression analysis using proper reference priors. Present a scatterplot of the data with the estimated regression line and a point-wise 95% probability band. Quantify the association between x and y using an approach of your choosing. Predict the value for new schools with $x = -16.04$. Present posterior inferences for parameters and predictions in an appropriately designed table. Would you be surprised if higher socioeconomic status were positively associated with higher test scores?

Table 9.7: *Coleman Report data.*

School	y	x	School	y	x
1	37.01	7.20	11	23.30	−12.86
2	26.51	−11.71	12	35.20	0.92
3	36.51	12.32	13	34.90	4.77
4	40.70	14.28	14	33.10	−0.96
5	37.10	6.31	15	22.70	−16.04
6	33.90	6.16	16	39.70	10.62
7	41.80	12.70	17	31.80	2.66
8	33.40	−0.17	18	31.70	−10.99

(e) Discuss a sensitivity analysis.

(f) Write-up of your entire analysis.

Dr. David Mannino, M.D. (Division of Pulmonary, Critical Care, and Sleep Medicine and Director of the Pulmonary Epidemiology Research Laboratory at the University of Kentucky) provided prior information. The values are measured in liters. The two numbers are the prior best guess of the mean FEV followed by the 99th percentile for the mean. For 18-year-old, female smokers, 70 inches tall: 4.0 and 4.8. For 16-year-old, male nonsmokers, 70 inches tall: 4.2 and 5.0. For 13-year-old, male smokers, 66 inches tall: 3.4 and 4.0. For 12-year-old, male nonsmokers, 60 inches tall: 2.7 and 3.5.

EXERCISE 9.22. *The Coleman Report Data.* Mosteller and Tukey (1977) and Christensen (1996) reproduced data collected from schools in the New England and Mid-Atlantic states of the USA. Consider two variables: y – the mean verbal test score for sixth graders and x – a composite measure of socioeconomic status associated with the school. The data are presented in Table 9.7. We wish to predict y based on x. Conduct a complete Bayesian regression analysis using proper reference priors. Present a scatterplot of the data with the estimated regression line and a point-wise 95% probability band. Quantify the association between x and y using an approach of your choosing. Predict the value for new schools with $x = -16.04$. Present posterior inferences for parameters and predictions in an appropriately designed table. Would you be surprised if higher socioeconomic status were positively associated with higher test scores?

Table 9.7: *Coleman Report data.*

School	y	x	School	y	x
1	37.01	7.20	11	23.30	−12.86
2	26.51	−11.71	12	35.20	0.92
3	36.51	12.32	13	34.90	4.77
4	40.70	14.28	14	33.10	−0.96
5	37.10	6.31	15	22.70	−16.04
6	33.90	6.16	16	39.70	10.62
7	41.80	12.70	17	31.80	2.66
8	33.40	−0.17	18	31.70	−10.99
9	41.01	9.85	19	43.10	15.03
10	37.20	−0.05	20	41.01	12.77

9.8 Nonlinear Regression*

Throughout this chapter we have assumed that $\mathrm{E}[y|x] = x'\beta$. We now allow a more general form for the regression function, $\mathrm{E}[y|x] = m(x;\beta)$ where $m(x;\beta)$ is a known function of x and β. In practice, x is observed but β is an r vector of unknown parameters. As before, although x is often random, we condition on it and so treat it as fixed. The nonlinear regression model is

$$\begin{aligned} y_i \mid \beta, \tau &\overset{ind}{\sim} N(\theta_i, 1/\tau) \\ \theta_i &= m(x_i;\beta). \end{aligned}$$

EXAMPLE 9.8.1. Carlin and Gelfand (1991) reported data from a growth study by Ratkowsky (1983) on length y and age x measurements collected on 27 dugongs, a large marine mammal. The data can be found at our website, along with R and WinBUGS programs for this example. Carlin and Gelfand consider a growth curve model that is similar to