Part3. Normal Model.

Elena Moltchanova

2021S2

A note on notation -1:

The normal distribution can be parametrised via mean and **standard deviation**:

$$X \sim N(\mu, \sigma)$$

or via mean and **variance** (the squared standard deviation):

$$X \sim N(\mu, \sigma^2)$$

A note on notation -2:

In Bayesian statistics, it is often convenient to express the normal distribution via **precision**, i.e., inverse variance:

$$X \sim N(\mu, \tau),$$

where $\tau = \frac{1}{\sigma^2}$.

Make sure to check when you are reading a new book or using a new software!

Normal distribution with known precision and unknown mean:

Let $y_1, ..., y_n$ be an i.i.d. sample from a normal distribution with unknown mean μ and known precision (inverse variance) τ:

$$y_i|\mu, \tau \sim N(\mu, \tau).$$

Assume a normal prior for μ:

$$\mu \sim N(\mu_0, \tau_0).$$

Analysis.

Let's derive the classical MLE for μ and the Bayesian posterior distribution for μ.

...

Normal distribution with unknown precision and known mean:

Let $y_1, ..., y_n$ be an i.i.d. sample from a normal distribution with unknown mean μ and known precision (inverse variance) τ:

$$y_i|\mu, \tau \sim N(\mu, \tau).$$

Assume now that the mean μ is known and it is the precision τ we are after. The precision (inverse variance) is a positive number, so a technically suitable distribution is a gamma distribution:

$$\tau \sim \Gamma(a, b).$$

Analysis.

Let's derive the classical MLE for τ and the Bayesian posterior distribution for τ.

...

Bayesian vs. Classical:

What happens to the posterior inference when the sample size n increases?

...

Introducing WinBUGS.

...

Summary of numerical methods:

- Use simulations (i.e., your own toy datasets) to learn new methods
- Don't forget to check for convergence

Linear Regression Model. Classical Set-Up:

$$y_i = a + bx_i + \epsilon_i$$

where

$$\epsilon_i \sim N(0, \sigma^2)$$

Note: σ^2 is the variance of the above normal distribution.

A note on interpretation:

$$y_i = a + bx_i + \epsilon_i$$

The intercept a is the expected value when $x = 0$.

Each one unit increase in X is **associated with an average** b unit increase/decrease in Y. (correlation does not imply causality!)

Sample-specific centering:

$$y_i = a + b(x_i - \bar{x}) + \epsilon_i$$

The intercept a is the expected value when $x = \bar{x}$.

Logging the response:

$$\log(y_i) = a + bx_i + \epsilon_i$$

Each one unit increase in X is **associated with an average** $(e^b - 1) \times 100\%$ increase/decrease in Y.

The linear regression is linear in parameters!!! (Not variables)

Linear Regression Model. Another Way:

$$y_i \sim N(\mu_i, \sigma^2)$$

where

$$\mu_i = a + bx_i$$

Note: σ^2 is the variance of the above normal distribution.

Linear Regression Model. Bayesian Way.

Likelihood:

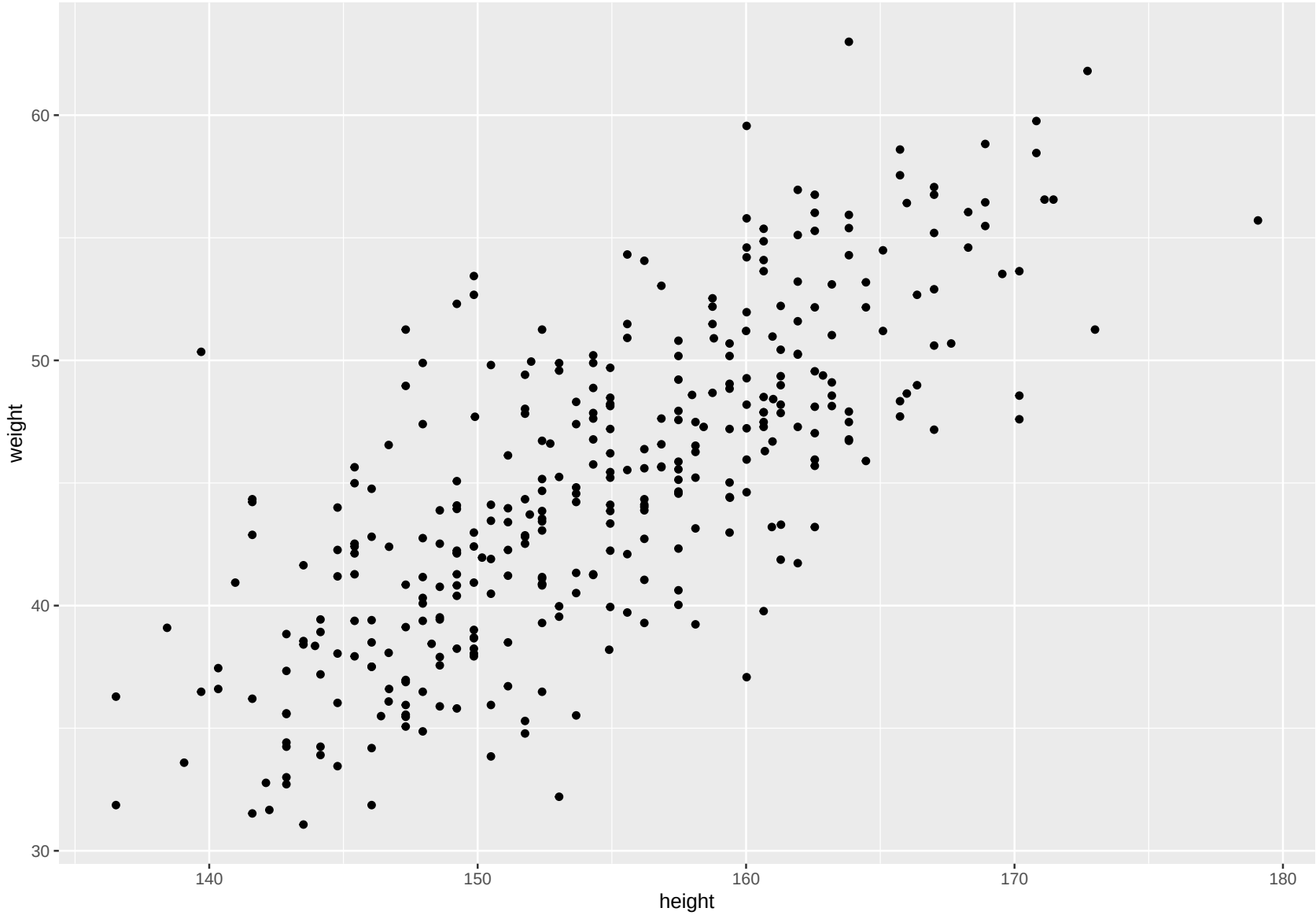
$$y_i|\mu_i,\tau \sim N(\mu_i,\tau) \quad (1)$$
$$\mu_i = a + bX \quad (2)$$

Priors:

$$a \sim N(\mu_a,\tau_a) \quad (3)$$
$$b \sim N(\mu_b,\tau_b) \quad (4)$$

$$\tau \sim Gamma(a_\tau,b_\tau) \quad (5)$$

Example: Howell's Data



Model:

$$\text{Weight}_i = a + b\text{Height}_i + \epsilon_i.$$

In Bayesian formulation:

$$\text{Weight}_i | \mu_i \sim N(\mu_i, \tau),$$

where

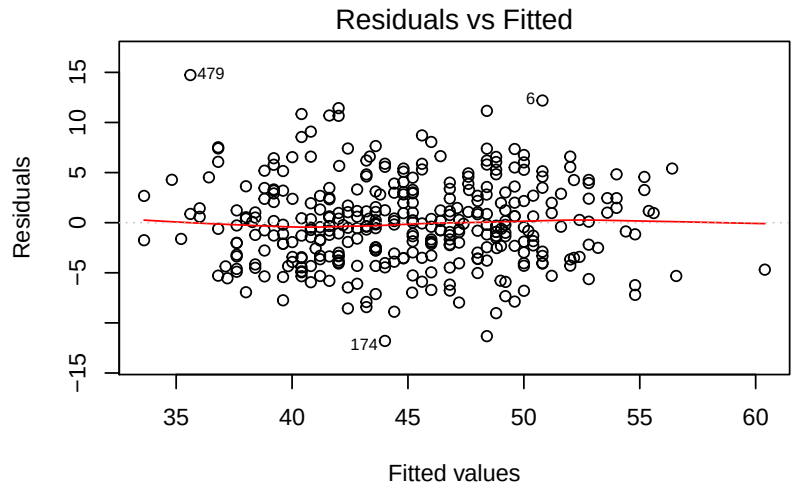
$$\mu_i = a + b\text{Height}_i.$$

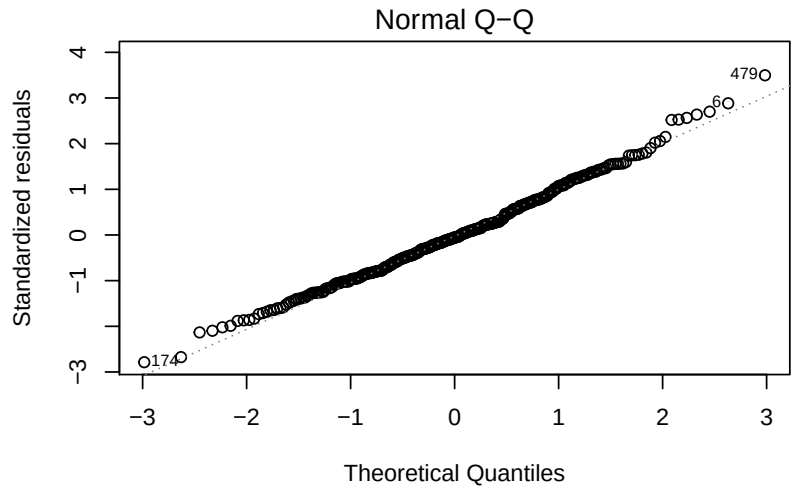
The classical way-1

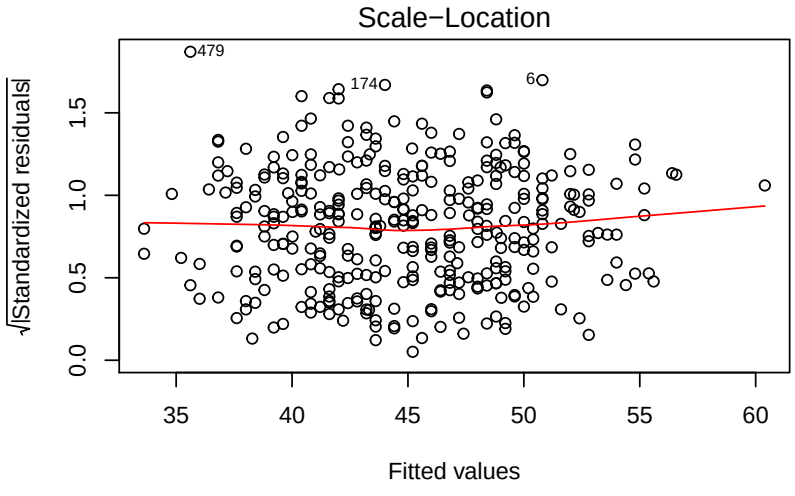
```
m1 <- lm(weight ~ height, data=d.adult)
summary(m1)
```

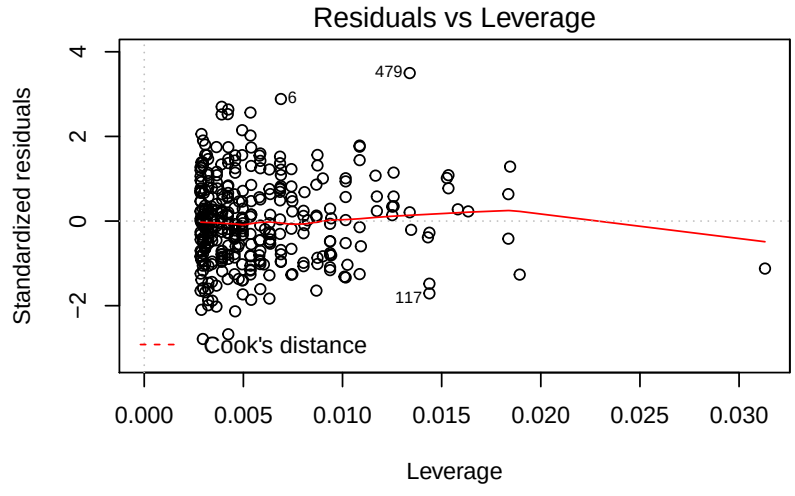
```
##
## Call:
## lm(formula = weight ~ height, data = d.adult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.8022  -3.0183  -0.2293   2.8117  14.7348
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -52.31618    4.52650  -11.56   <2e-16 ***
## height        0.62942    0.02924   21.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.242 on 350 degrees of freedom
## Multiple R-squared:  0.5696, Adjusted R-squared:  0.5684
## F-statistic: 463.3 on 1 and 350 DF,  p-value: < 2.2e-16
```

Classical assumption checking









The classical way-2

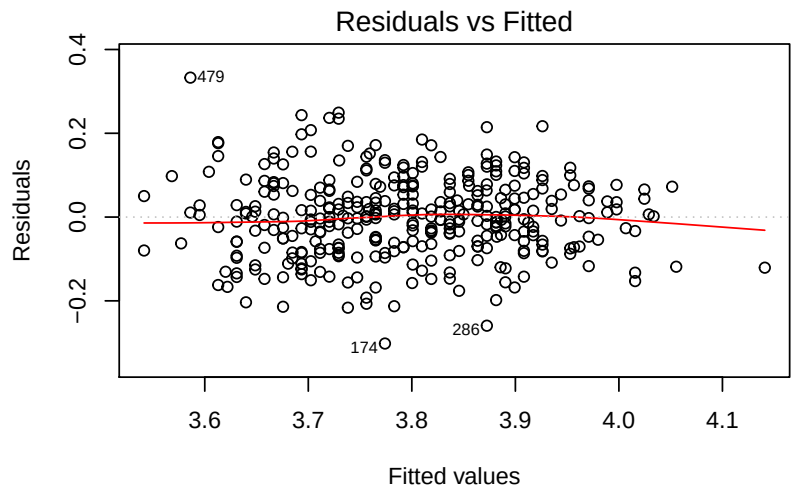
```r
m2 <- lm(weight ~ I(height-mean(height)), data=d.adult)
summary(m2)
```

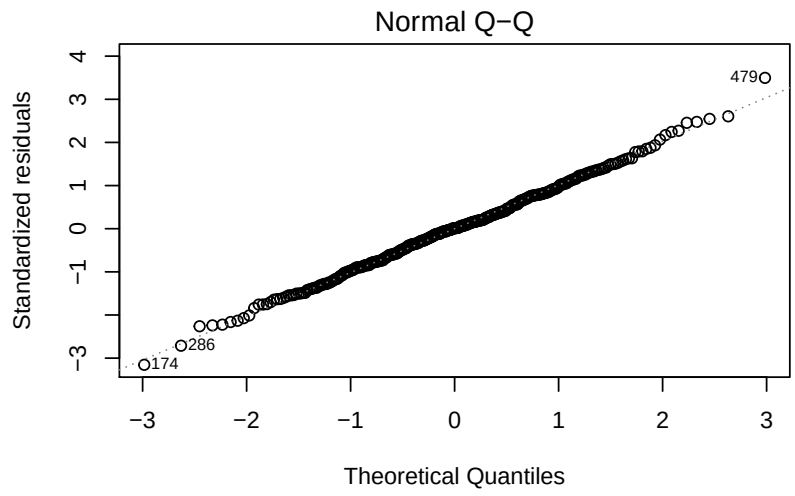
```
##
## Call:
## lm(formula = weight ~ I(height - mean(height)), data = d.adult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.8022  -3.0183  -0.2293   2.8117  14.7348
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                44.99049    0.22609  199.00   <2e-16 ***
## I(height - mean(height))    0.62942    0.02924   21.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.242 on 350 degrees of freedom
## Multiple R-squared:  0.5696, Adjusted R-squared:  0.5684
## F-statistic: 463.3 on 1 and 350 DF,  p-value: < 2.2e-16
```

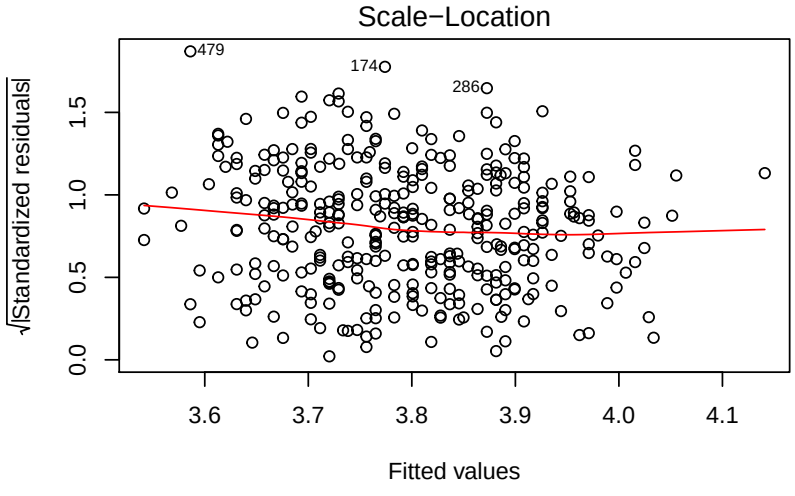
The classical way-3

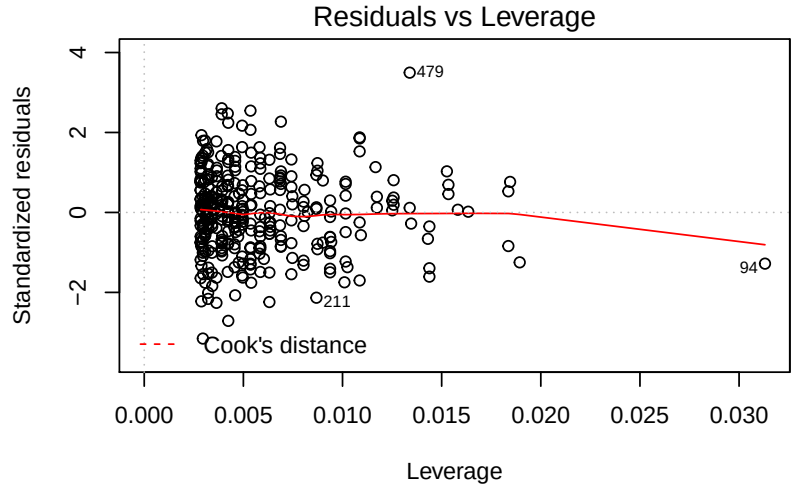
```
m3 <- lm(log(weight) ~ height, data=d.adult)
summary(m3)
```

```
##
## Call:
## lm(formula = log(weight) ~ height, data = d.adult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30191 -0.06556  0.00108  0.06525  0.33286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6174274  0.1022805   15.81   <2e-16 ***
## height      0.0140923  0.0006608   21.33   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09585 on 350 degrees of freedom
## Multiple R-squared:  0.5651, Adjusted R-squared:  0.5639
## F-statistic: 454.8 on 1 and 350 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
479
174
286
0.4
0.2
0.0
−0.2
3.6
3.7
3.8
3.9
4.0
4.1


Normal Q−Q
Standardized residuals
Theoretical Quantiles
479
286
174
4
3
2
1
0
−1
−3
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
Fitted values
479
174
286
1.5
1.0
0.5
0.0
3.6
3.7
3.8
3.9
4.0
4.1


Residuals vs Leverage
Standardized residuals
Leverage
479
211
94
Cook's distance
4
2
0
−2
0.000
0.005
0.010
0.015
0.020
0.025
0.030

The Bayesian way

```
library(MCMCglmm)

m1b <- MCMCglmm(weight ~ height, data=d.adult, verbose=F)
```

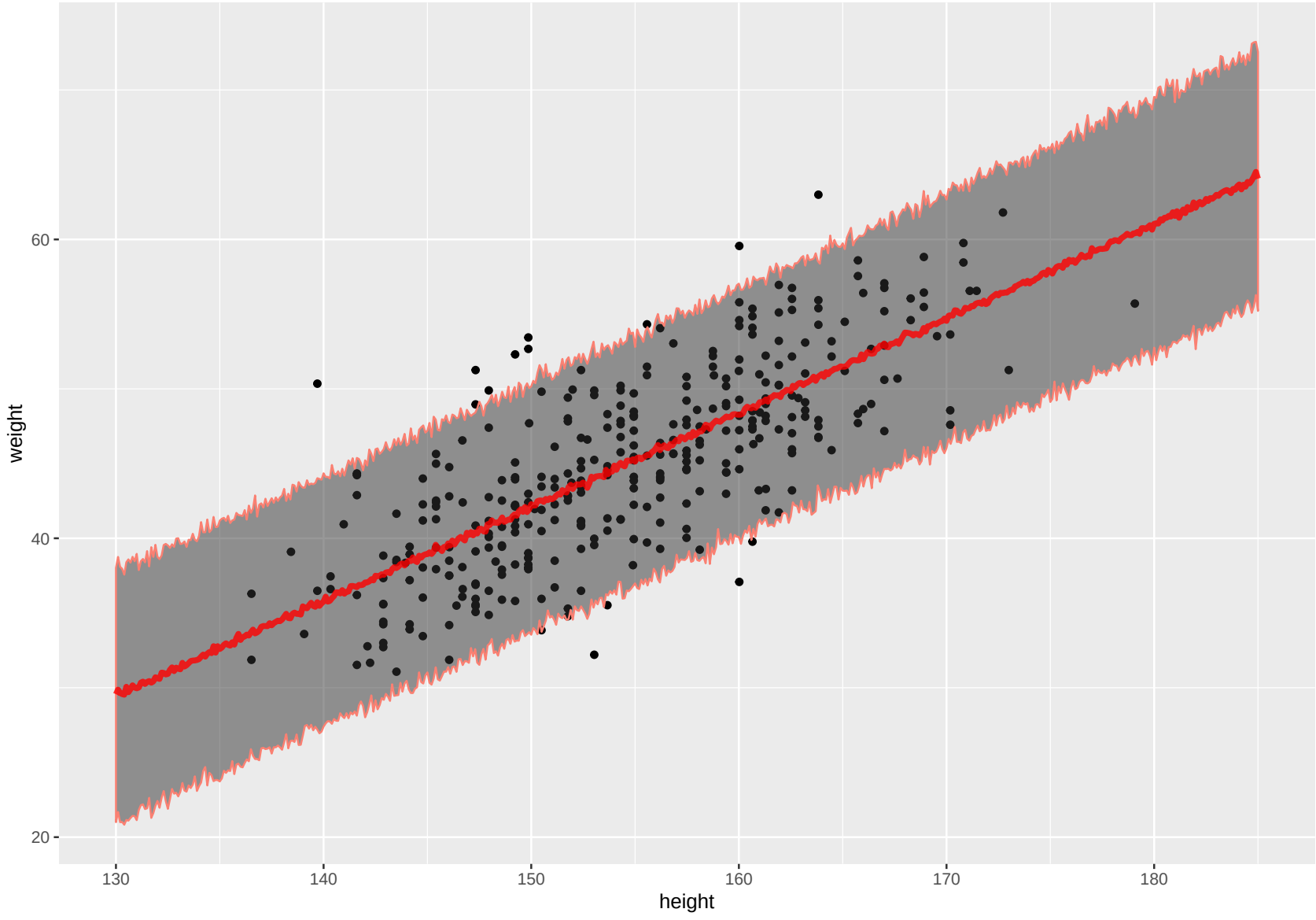
```r
summary(m1b)
```

```
##
##  Iterations = 3001:12991
##  Thinning interval  = 10
##  Sample size  = 1000
##
##  DIC: 2020.144
##
##  R-structure:  ~units
##
##       post.mean l-95% CI u-95% CI eff.samp
## units     18.12    15.47    21.04    899.6
##
##  Location effects: weight ~ height
##
##             post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)  -52.4646 -60.7339 -43.8509     1000 <0.001 ***
## height         0.6305   0.5742   0.6844     1000 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bayesian Diagnostics: Model Fit

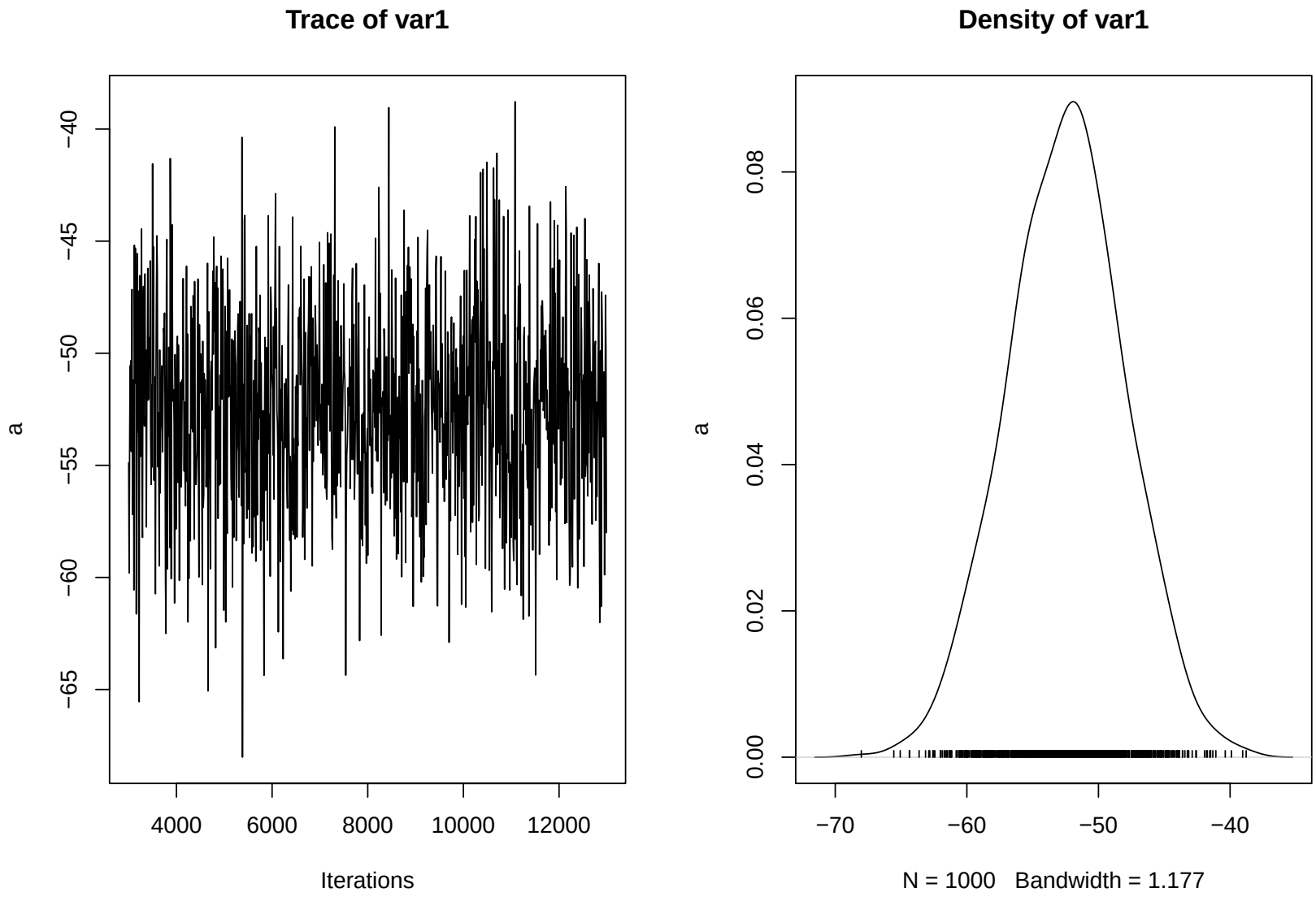
```
dtf <- data.frame(height=seq(130,185,.1))

# a function to produce an posterior pred. estimate
pred.reg <- function(x,a,b,sigma){
  a+b*x+rnorm(length(a),0,sd=sigma)
}

m1b.pred <- sapply(dtf$height, pred.reg,
      a=m1b$Sol[,1],b=m1b$Sol[,2],sigma=sqrt(m1b$VCV))

dtf$pp.mean <- apply(m1b.pred,2,mean)
dtf$pp.q025 <- apply(m1b.pred,2,quantile,.025)
dtf$pp.q975 <- apply(m1b.pred,2,quantile,.975)
```

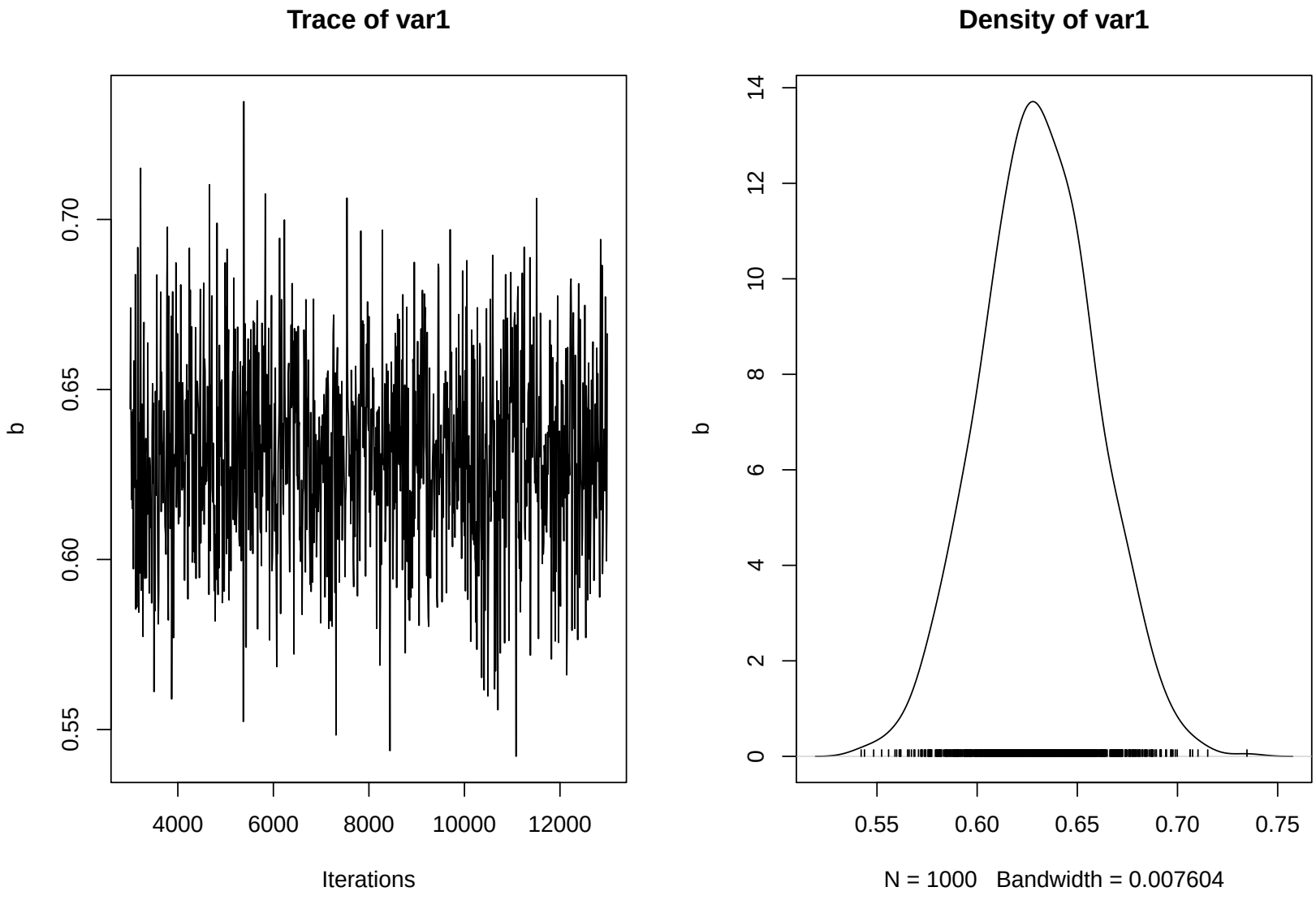
Plotting



Bayesian diagnostics: convergence of *a*

```
plot(m1b$Sol[,1],ylab='a')
```



Bayesian diagnostics: convergence of *b*

```
plot(m1b$Sol[,2],ylab='b')
```