Reading comprehension test Analysis

Matteo Pedone

12/03/21

Reading Comprehension test data

The following example is reported in Hoff (2009). It is the *Reading comprehension* example reported in chapter 7 *"The multivariate normal model"*.

```
Y <- reading
n <- dim(Y)[1]
p <- dim(Y)[2]
n
```

```
## [1] 22
```

```
p
```

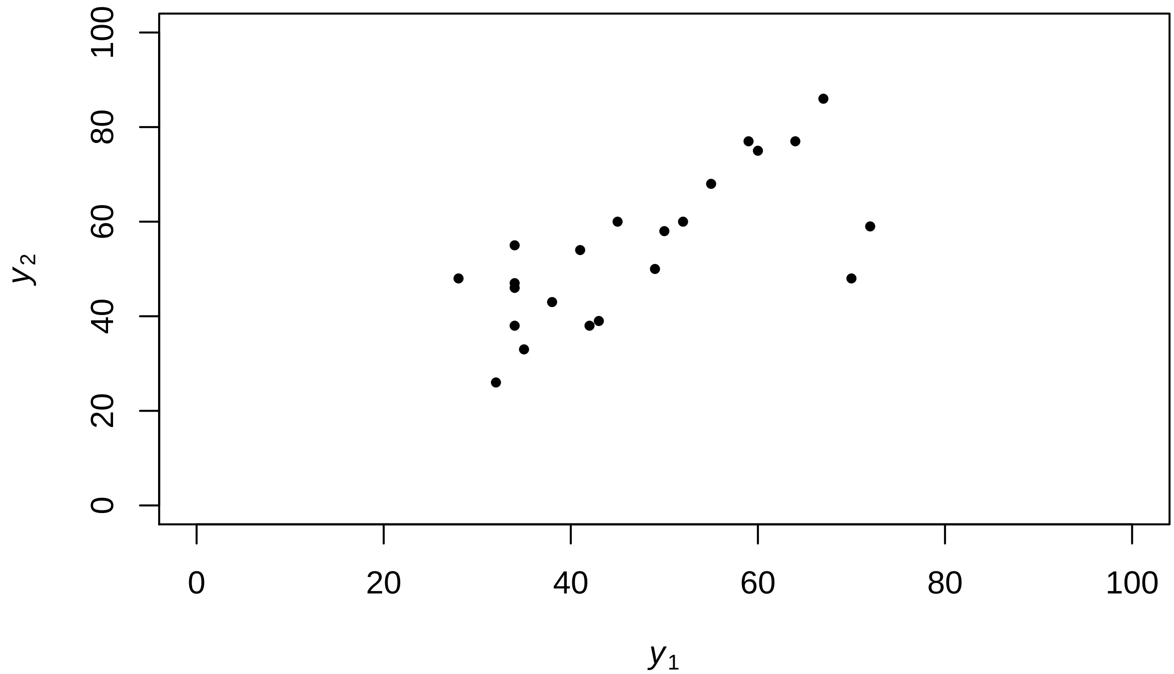
```
## [1] 2
```

```
summy <- summary(Y)
summy
```

```
##     pretest         posttest    
##  Min.   :28.00   Min.   :26.00  
##  1st Qu.:34.25   1st Qu.:43.75  
##  Median :44.00   Median :52.00  
##  Mean   :47.18   Mean   :53.86  
##  3rd Qu.:58.00   3rd Qu.:60.00  
##  Max.   :72.00   Max.   :86.00  
```

Descriptives

The observed values $y_1, \ldots, y_{22}$ are plotted in the following figure.



The sample mean is $\bar{\boldsymbol{y}} = (47.18, 53.86)^T$, the sample variances are $s_{21} = 182.16$ and $s_{22} = 243.65$, and the sample correlation is $s_{1,2}/(s_1 \times s_2) = 0.7$.

Multivariate Normal Model

We will model these 22 pairs of scores as i.i.d. samples from a multivariate normal distribution. We consider a 2-dimensional data vector $\boldsymbol{y}$, its sampling density is given by

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \Sigma) = (2\pi)^{-p/2} \mid \Sigma \mid^{-1/2} \exp\{-(\boldsymbol{y} - \boldsymbol{\theta})^T \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\theta})/2\}.$$

Prior distributions

The prior distributions are:

$$p(\boldsymbol{\theta}) = MVN(\boldsymbol{\mu_0}, \boldsymbol{\Lambda_0}),$$

$$p(\Sigma) = IW(\boldsymbol{S}_0, \nu_0).$$

Following Hoff (2009) 's guidelines our prior expectation is $\boldsymbol{\mu}_0 = (50, 50)^T$ and prior covariance

$$\boldsymbol{\Lambda}_0 = \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix}.$$

As for the prior distribution on $\boldsymbol{\Sigma}$, we'll take $\boldsymbol{S}_0$ to be the same as $\boldsymbol{\Lambda}_0$, but only loosely center $\boldsymbol{\Sigma}$ around this value by taking $\nu_0 = p + 2$.

Posterior Computation

Let's use the Gibbs sampler to sample from the full conditional distributions. In this way we obtain an MCMC approximation to the joint posterior distribution $p(\boldsymbol{\theta}, \Sigma \mid y_1, \ldots, y_n)$. Combining sample informations with our prior distributions we obtain estimates and confidence intervals for the population parameters.

```
THETA <- SIGMA <- NULL
YS <- NULL
S <- 5000
cat("here starts the mcmc!", "\n")
```

```
## here starts the mcmc!
```

```r
for (s in 1:S) {
  ### update theta
  Ln <- solve(solve(L0) + n * solve(Sigma))
  mun <- Ln %*% (solve(L0) %*% mu0 + n * solve(Sigma) %*% ybar)
  theta <- rmvnorm(1, mun, Ln)
  ### update Sigma
  Sn <- solve(S0 + (t(Y) - c(theta)) %*% t(t(Y) - c(theta)))
  Sigma <- solve(rwish(1, nu0 + n, Sn))
  ###
  YS <- rbind(YS, rmvnorm(1, theta, Sigma))
  ### save results
  THETA <- rbind(THETA, theta)
  SIGMA <- rbind(SIGMA, c(Sigma))
}
cat("sampling is done", "\n")
```

```
## sampling is done
```

The above code generates 5000 values $(\{\boldsymbol{\theta}^{(1)}, \Sigma^{(1)}\}), \ldots, (\{\boldsymbol{\theta}^{(5000)}, \Sigma^{(5000)}\})$, whose empirical distribution approximates $p(\boldsymbol{\theta}, \Sigma \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

From these samples we can approximate posterior probabilities and confidence regions of interest.

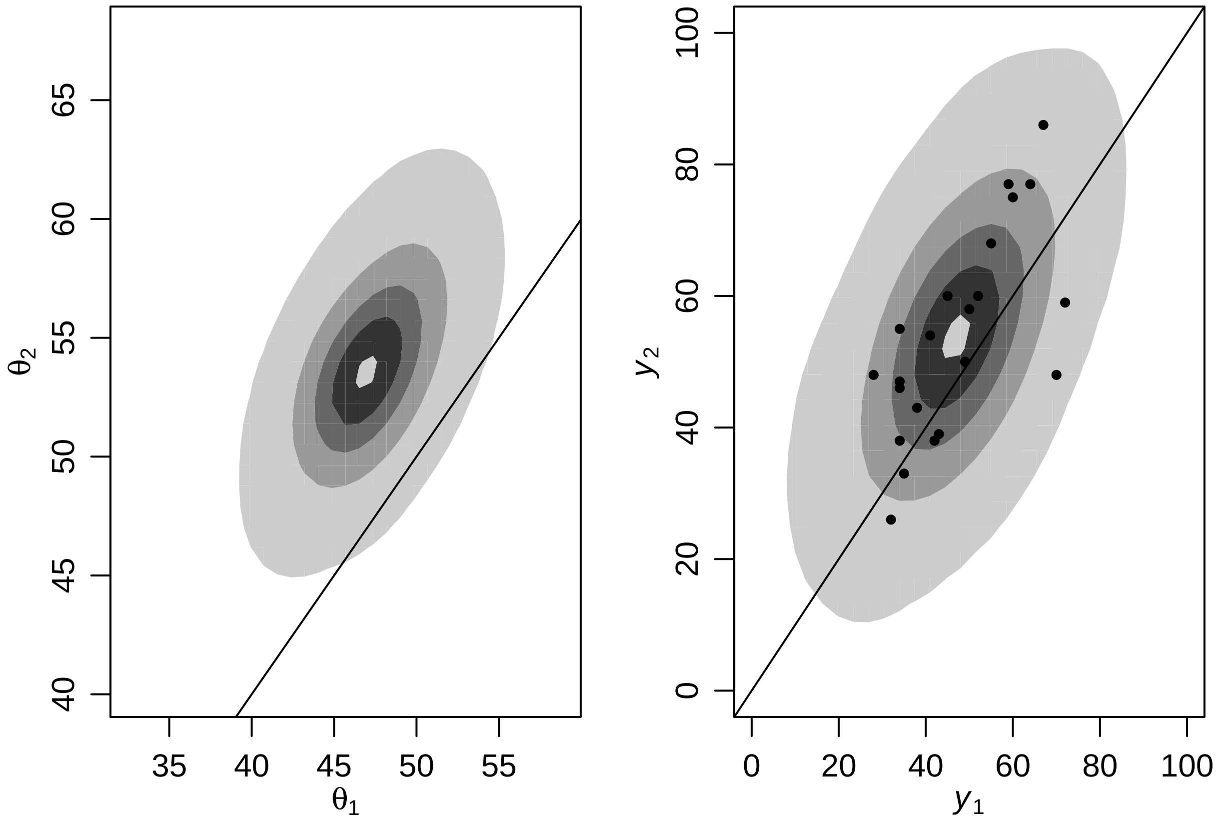
```r
#load("../data/res.RData")
#THETA <- output$Thetapost
#SIGMA <- output$Sigmapost
#YS <- output$Ypost

quantile(THETA[, 2] - THETA[, 1], prob = c(.025, .5, .975))
```

```
##      2.5%       50%     97.5%
##  1.480852  6.610902 11.679931
```

Conclusion

The posterior probability $Pr(\theta_2 > \theta_1 \mid y1, \ldots, y_n) = 0.993$ indicates strong evidence that, if we were to give exams and instruction to a large population of children, then the average score on the second exam would be higher than that on the first. This evidence is displayed graphically in the first panel of next figure, which shows $97.5\%, 75\%, 50\%, 25\%$ and 2.5% highest posterior density contours for the joint posterior distribution of $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$. A highest posterior density contour is a two-dimensional analogue of a confidence interval. The contours for the posterior distribution of $\boldsymbol{\theta}$ are all mostly above the $45-$degree line $\theta_1 = \theta_2$.



References

Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods*. Vol. 580. Springer.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. CRC Press.