ESC 2022 WEEK5 Hierarchical Model-Part1 Group comparisons

학술부- 김송희, 김수민, 선대운

August 11, 2022

8.1 Comparing two groups

8.2 Comparing multiple groups

We parameterize the two population means by their average and their difference.

This type of parameterization is extended to the multigroup case, where the average group mean and the differences across group means are described by a normal sampling model.

Introduction to Hierarchical Model

8.1 Comparing two groups

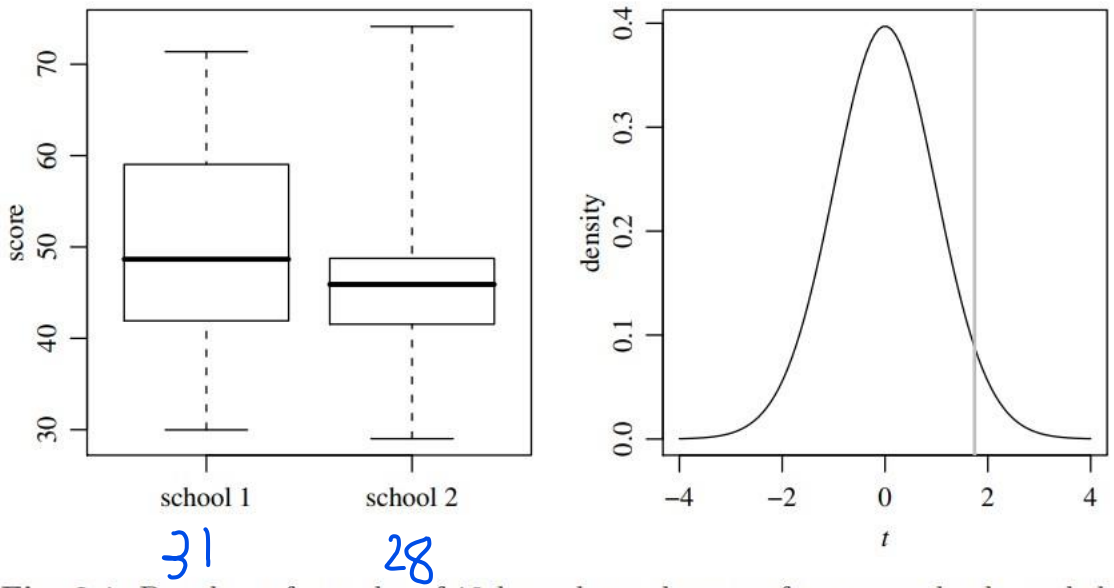


Fig. 8.1. Boxplots of samples of 10th grade math scores from two schools, and the null distribution for testing equality of the population means. The gray line indicates the observed value of the t-statistic.

$\theta 1$ - average score if all 10th graders in school1 were tested

$\theta 2$ - average score if all 10th graders in school2 were tested

$\bar{y}_1 = 50.81$, $\bar{y}_2 = 46.15$ → $\theta 1 > \theta 2$? → t-statistic

4.66

8.1 Comparing two groups

$$t(\boldsymbol{y}_1, \boldsymbol{y}_2) = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{50.81 - 46.15}{10.44\sqrt{\frac{1}{31}+\frac{1}{28}}} = 1.74$$

$n_1 + n_2 - 2 = 57$

$$s_p{}^2 = \frac{[(n_1-1)s_1{}^2+(n_2-1)s_2{}^2]}{(n_1+n_2-2)}$$

Pooled estimate of the population variance

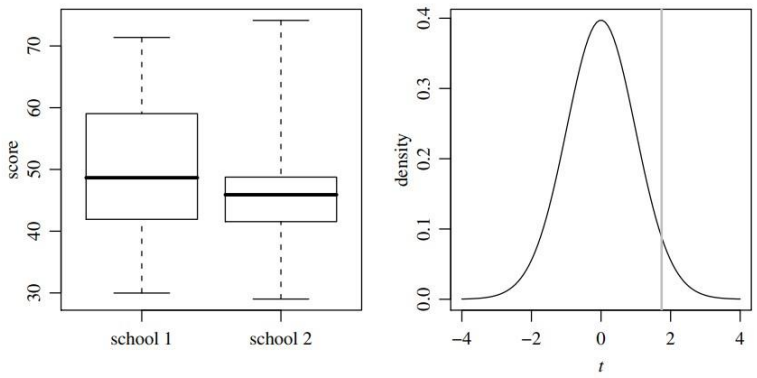


Fig. 8.1. Boxplots of samples of 10th grade math scores from two schools, and the null distribution for testing equality of the population means. The gray line indicates the observed value of the t-statistic.

p = 0.087

8.1 Comparing two groups

p-value에 기반한 모델 선택

If $p < 0.05$,
- reject the model that the two groups have the same distribution;
- conclude that $\theta_1 \neq \theta_2$;
- use the estimates $\hat{\theta}_1 = \bar{y}_1$, $\hat{\theta}_2 = \bar{y}_2$.

If $p > 0.05$
- accept the model that the two groups have the same distribution;
- conclude that $\theta_1 = \theta_2$;
- use the estimates $\hat{\theta}_1 = \hat{\theta}_2 = (\sum y_{i,1} + \sum y_{i,2})/(n_1 + n_2)$.

Completely distinct or exactly indentical?

$$\widehat{\theta_1} = w\overline{y_1} + (1 - w)\,\overline{y_2}$$

$$w = 1 \quad if\ p < 0.05$$

$$w = \frac{n_1}{n_1 + n_2} \quad if\ \ p > 0.05$$

8.1 Comparing two groups

Allow w to vary continuously and have a value that depends on such things as the relative sample sizes n_1, n_2, the sampling variablility σ^2 and our prior information about the similarities of the two populations.

$$
\begin{aligned}
Y_{i,1} &= \mu + \delta + \epsilon_{i,1} \\
Y_{i,2} &= \mu - \delta + \epsilon_{i,2} \\
\{\epsilon_{i,j}\} &\sim \text{i.i.d. normal}(0, \sigma^2)
\end{aligned}
$$

$$
\mu = \frac{\theta_1 + \theta_2}{2}
$$

$$
\begin{aligned}
\theta_1 &= \mu + \delta \\
\theta_2 &= \mu - \delta
\end{aligned}
$$

$$
\delta = \frac{\theta_1 - \theta_2}{2}
$$

8.1 Comparing two groups

$$
\begin{aligned}
p(\mu, \delta, \sigma^2) &= p(\mu) \times p(\delta) \times p(\sigma^2) \\
\mu &\sim \text{normal}(\mu_0, \gamma_0^2) \\
\delta &\sim \text{normal}(\delta_0, \tau_0^2) \\
\sigma^2 &\sim \text{inverse-gamma}(\nu_0/2, \nu_0\sigma_0^2/2).
\end{aligned}
$$

$$
\begin{aligned}
&\{\mu|\boldsymbol{y}_1, \boldsymbol{y}_2, \delta, \sigma^2\} \sim \text{normal}(\mu_n, \gamma_n^2), \text{ where} \\
&\quad \mu_n = \gamma_n^2 \times [\mu_0/\gamma_0^2 + \textstyle\sum_{i=1}^{n_1}(y_{i,1} - \delta)/\sigma^2 + \sum_{i=1}^{n_2}(y_{i,2} + \delta)/\sigma^2] \\
&\quad \gamma_n^2 = [1/\gamma_0^2 + (n_1 + n_2)/\sigma^2]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&\{\delta|\boldsymbol{y}_1, \boldsymbol{y}_2, \mu, \sigma^2\} \sim \text{normal}(\delta_n, \tau_n^2), \text{ where} \\
&\quad \delta_n = \tau_n^2 \times [\delta_0/\tau_0^2 + \textstyle\sum(y_{i,1} - \mu)/\sigma^2 - \sum(y_{i,2} - \mu)/\sigma^2] \\
&\quad \tau_n^2 = [1/\tau_0^2 + (n_1 + n_2)/\sigma^2]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&\{\sigma^2|\boldsymbol{y}_1, \boldsymbol{y}_2, \mu, \delta\} \sim \text{inverse-gamma}(\nu_n/2, \nu_n\sigma_n^2/2), \text{ where} \\
&\quad \nu_n = \nu_0 + n_1 + n_2 \\
&\quad \nu_n\sigma_n^2 = \nu_0\sigma_0^2 + \textstyle\sum(y_{i,1} - [\mu + \delta])^2 + \sum(y_{i,2} - [\mu - \delta])^2
\end{aligned}
$$

8.1 Comparing two groups

$V_0 = 0$

$$\sigma_n^2 = \frac{\sum(y_{i,1} - [\mu + \delta])^2 + \sum(y_{i,2} - [\mu - \delta])^2}{n_1 + n_2}$$

$$\mu_n = \frac{\sum(y_{i,1} - \delta) + \sum(y_{i,2} + \delta)}{n_1 + n_2}, \quad \delta_n = \frac{\sum(y_{i,1} - \mu) - \sum(y_{i,2} - \mu)}{n_1 + n_2}$$

$$\overline{y_1} = \mu_n + \delta_n$$
$$\overline{y_2} = \mu_n - \delta_n$$

$m_0 = \delta_0 = 0$

$\tau_0^2 = T_0^2 = \infty$

8.1 Comparing two groups

Analysis of the math score data

$m_0 = 50,\ {\sigma_0}^2 = 10^2 = 100$

${\gamma_0}^2 = 25^2 = 625,\ \ \nu_0 = 1,\ \ \delta_0 = 0\ \ ,\ \ {\tau_0}^2 = 25^2 = 625$

$$p(\mu, \delta, \sigma^2) = p(\mu) \times p(\delta) \times p(\sigma^2)$$
$$\mu \sim \text{normal}(\mu_0, \gamma_0^2)$$
$$\delta \sim \text{normal}(\delta_0, \tau_0^2)$$
$$\sigma^2 \sim \text{inverse-gamma}(\nu_0/2, \nu_0\sigma_0^2/2).$$

$\{\mu|\boldsymbol{y}_1, \boldsymbol{y}_2, \delta, \sigma^2\} \sim \text{normal}(\mu_n, \gamma_n^2)$, where
$\mu_n = \gamma_n^2 \times [\mu_0/\gamma_0^2 + \sum_{i=1}^{n_1}(y_{i,1} - \delta)/\sigma^2 + \sum_{i=1}^{n_2}(y_{i,2} + \delta)/\sigma^2]$
$\gamma_n^2 = [1/\gamma_0^2 + (n_1 + n_2)/\sigma^2]^{-1}$

$\{\delta|\boldsymbol{y}_1, \boldsymbol{y}_2, \mu, \sigma^2\} \sim \text{normal}(\delta_n, \tau_n^2)$, where
$\delta_n = \tau_n^2 \times [\delta_0/\tau_0^2 + \sum(y_{i,1} - \mu)/\sigma^2 - \sum(y_{i,2} - \mu)/\sigma^2]$
$\tau_n^2 = [1/\tau_0^2 + (n_1 + n_2)/\sigma^2]^{-1}$

$\{\sigma^2|\boldsymbol{y}_1, \boldsymbol{y}_2, \mu, \delta\} \sim \text{inverse-gamma}(\nu_n/2, \nu_n\sigma_n^2/2)$, where
$\nu_n = \nu_0 + n_1 + n_2$
$\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + \sum(y_{i,1} - [\mu + \delta])^2 + \sum(y_{i,2} - [\mu - \delta])^2$

Initialize: pick arbitrary starting value $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \cdots, \theta_i^{(1)}, \theta_{i+1}^{(1)}, \cdots, \theta_K^{(1)})$

Iterate a Cycle:

Step 1. draw $\theta_1^{(s+1)} \sim \pi(\theta_1|\theta_2^{(s)}, \theta_3^{(s)}, \cdots, \theta_K^{(s)}, y)$

Step 2. draw $\theta_2^{(s+1)} \sim \pi(\theta_2|\theta_1^{(s+1)}, \theta_3^{(s)}, \cdots, \theta_K^{(s)}, y)$

$\vdots$

Step i. draw $\theta_i^{(s+1)} \sim \pi(\theta_i|\theta_1^{(s+1)}, \theta_2^{(s+1)}, \cdots, \theta_{i-1}^{(s+1)}, \theta_{i+1}^{(s)}, \cdots, \theta_K^{(s)}, y)$

Step i+1. draw $\theta_{i+1}^{(s+1)} \sim \pi(\theta_{i+1}|\theta_1^{(s+1)}, \theta_2^{(s+1)}, \cdots, \theta_i^{(s+1)}, \theta_{i+2}^{(s)}, \cdots, \theta_K^{(s)}, y)$

$\vdots$

Step K. draw $\theta_K^{(s+1)} \sim \pi(\theta_K|\theta_1^{(s+1)}, \theta_2^{(s+1)}, \cdots, \theta_{K-1}^{(s+1)}, y)$

8.1 Comparing two groups

```
[ ]  import numpy as np
     import pandas as pd
     import random

     y1 = np.array([52.11, 57.65, 66.44, 44.68, 40.57, 35.04, 50.71, 66.17, 39.43,
     46.17, 58.76, 47.97, 39.18, 64.63, 69.38, 32.38, 29.98, 59.32,
     43.04, 57.83, 46.07, 47.74, 48.66, 40.8, 66.32, 53.7, 52.42,
     71.38, 59.66, 47.52, 39.51])
     n1 = len(y1)
```

```
[ ]

     y2 = np.array([52.87, 50.03, 41.51, 37.42, 64.42, 45.44, 46.06, 46.37, 46.66,
     29.01, 35.69, 49.16, 55.9, 45.84, 35.44, 43.21, 48.36, 74.14,
     46.76, 36.97, 43.84, 43.24, 56.9, 47.64, 38.84, 42.96, 41.58,
     45.96])

     n2 = len(y2)
```

```
     ### 앞에서 설정한 사전 파라미터

     mu0 = 50 ; g02 = 625
     delta0 = 0 ; t02 = 625
     s20 = 100 ; nu0 = 1
```

```
[ ]  ### 시작 값
     mu = (np.mean(y1) + np.mean(y2)) / 2
     delta = (np.mean(y1) - np.mean(y2)) / 2
```

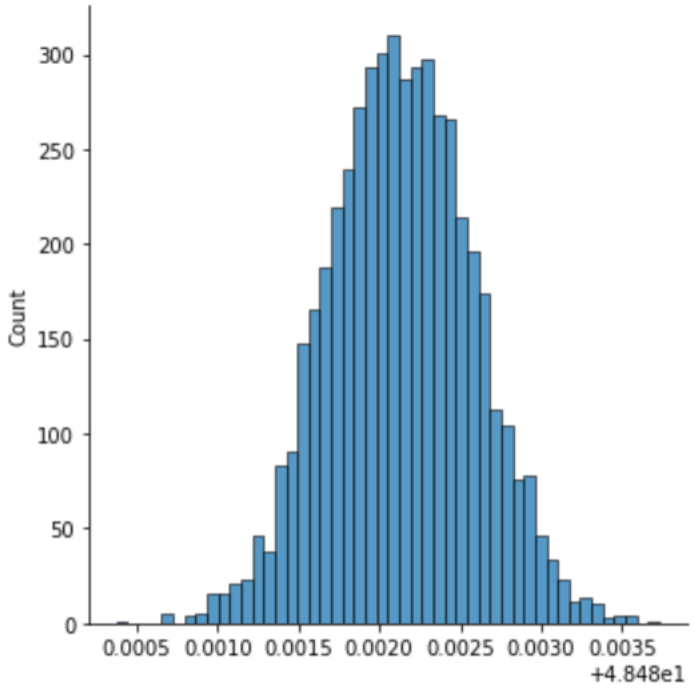
8.1 Comparing two groups

```
MU = list()
DEL = list()
S2 = list()
random.seed(1)

for s in np.arange(0,5000):

    ## s2 업데이트
    s2 = 1/np.random.gamma((nu0 + n1 + n2)/2, ( nu0 * s20 + np.sum((y1 - mu - delta)**2) + np.sum((y2 - mu + delta)**2) )/2 ,1)
    ##
    ## mu 업데이트
    var_mu = 1 / (1/g02 + (n1+n2) / s2)
    mean_mu = var_mu * (mu0 / g02 + sum(y1 - delta)/s2 + sum(y2+delta)/s2)
    mu = np.random.normal(mean_mu, np.sqrt(var_mu), 1)
    ##
    ## del 업데이트
    var_del = 1 / (1/t02 + (n1+n2)/s2)
    mean_del = var_del * (delta0/t02 + sum(y1 - mu)/s2 - sum(y2-mu)/s2)
    delta = np.random.normal(mean_del, np.sqrt(var_del), 1)
    ##
    ## 파라미터 값 저장
    MU = np.append(MU, mu)
    DEL = np.append(DEL, delta)
    S2 = np.append(S2,s2)
```

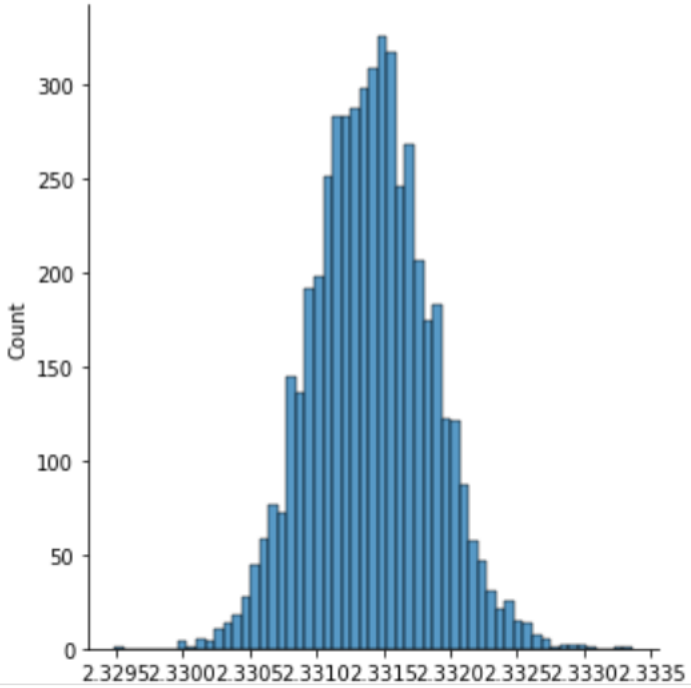
8.1 Comparing two groups

```
import seaborn as sns

sns.displot(MU)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f4a4c027e50>
```




```
import seaborn as sns

g= sns.displot(DEL)
```



8.1 Comparing two groups

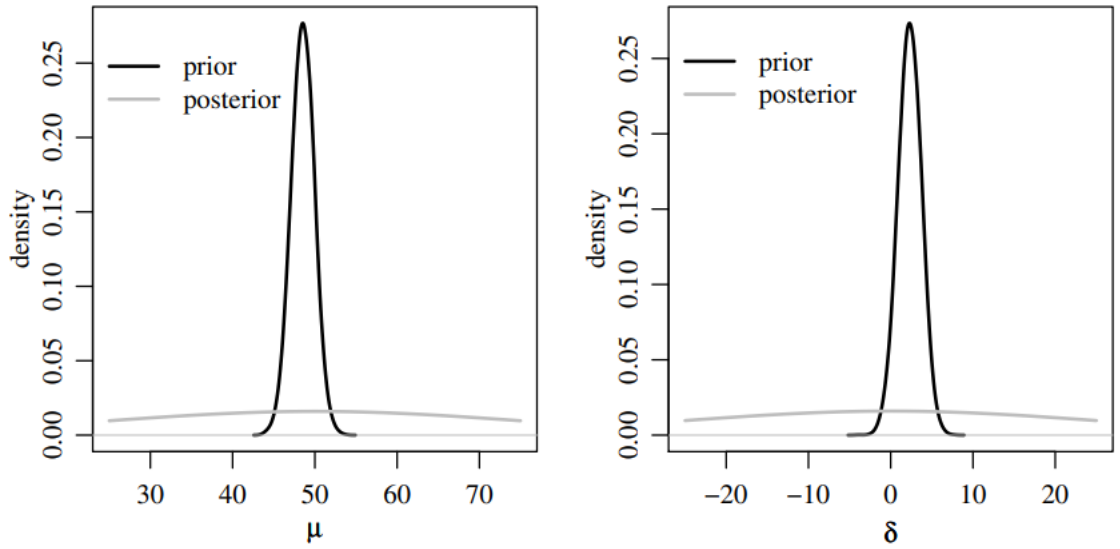


Fig. 8.2. Prior and posterior distributions for μ and δ.

$2\delta = (-0.61, 9.98)$

$Pr(\theta_1 > \theta_2 | \boldsymbol{y}_1, \boldsymbol{y}_2) = \Pr(\delta > 0 | \boldsymbol{y}_1, \boldsymbol{y}_2) \approx 0.96$

$pr(\delta > 0) = 0.5$

$pr(Y_1 > Y_2 | \boldsymbol{y}_1, \boldsymbol{y}_2) \approx 0.62$

8.2 Comparing multiple groups

Hierarchical or multilevel

patients within several hospitals

Genes within a group of animals

People within counties within regions within countries

$$y_{i,j}$$

8.2 Comparing multiple groups

Exchangeability and hierachical models

Let $p(y_1, \ldots, y_n)$ be the joint density of $Y_1, \ldots, Y_n$.
If $p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n})$
for all permutations π of $\{1, \ldots, n\}$, then $Y_1, \ldots, Y_n$ are exchangeable.

2-8. de Finetti's theorem

1) Conditional IID ⟹ exchangeability (곱셈)
2) Conditional IID ⟸ infinite exchangeability : de finetti's therem

$$p(y_1, \ldots, y_n) = \int \left\{ \prod_i p(y_i|\theta) \right\} p(\theta)\, d\theta$$

⟹ decomposition of the model

$$\phi \sim p(\phi)$$
$$\{Y_1, \ldots, Y_n|\phi\} \sim \text{i.i.d. } p(y|\phi).$$

8.2 Comparing multiple groups

$p(\boldsymbol{y}_j) = p(y_{1,j}, \ldots, y_{n,j})$

independent =>

$p(y_{n_j,j} \mid y_{1,j} \cdots y_{n_j-1,j}) = p(y_{n_j,j})$

De Finetti 정리 + diaconis and Freedman(1980)의 결과 ->

$$\{Y_{1,j}, \ldots, Y_{n_j,j} | \phi_j\} \sim \text{ i.i.d. } p(y|\phi_j)\,.$$

$$\{\phi_1, \ldots, \phi_m | \psi\} \sim \text{ i.i.d. } p(\phi|\psi)$$

8.2 Comparing multiple groups

$$
\begin{aligned}
\{y_{1,j}, \ldots, y_{n_j,j} | \phi_j\} &\sim \text{ i.i.d. } p(y|\phi_j) && \text{(within-group sampling variability)} \\
\{\phi_1, \ldots, \phi_m | \psi\} &\sim \text{ i.i.d. } p(\phi|\psi) && \text{(between-group sampling variability)} \\
\psi &\sim p(\psi) && \text{(prior distribution)}
\end{aligned}
$$

8.2 Comparing multiple groups

Framework [edit]

Let y_j be an observation and θ_j a parameter governing the data generating process for y_j. Assume further that the parameters $\theta_1, \theta_2, \ldots, \theta_j$ are generated exchangeably from a common population, with distribution governed by a hyperparameter ϕ.
The Bayesian hierarchical model contains the following stages:

$\text{Stage I: } y_j \mid \theta_j, \phi \sim P(y_j \mid \theta_j, \phi)$
$\text{Stage II: } \theta_j \mid \phi \sim P(\theta_j \mid \phi)$
$\text{Stage III: } \phi \sim P(\phi)$

The likelihood, as seen in stage I is $P(y_j \mid \theta_j, \phi)$, with $P(\theta_j, \phi)$ as its prior distribution. Note that the likelihood depends on ϕ only through θ_j.

The prior distribution from stage I can be broken down into:

$P(\theta_j, \phi) = P(\theta_j \mid \phi)P(\phi)$ *[from the definition of conditional probability]*

With ϕ as its hyperparameter with hyperprior distribution, $P(\phi)$.

Thus, the posterior distribution is proportional to:

$P(\phi, \theta_j \mid y) \propto P(y_j \mid \theta_j, \phi)P(\theta_j, \phi)$ *[using Bayes' Theorem]*
$P(\phi, \theta_j \mid y) \propto P(y_j \mid \theta_j)P(\theta_j \mid \phi)P(\phi)$[13]

끝

Part2 Contents

8.3

The hierarchical normal model

Hierarchical normal model

A popular model for describing the heterogeneity of means across several populations is the hierarchical normal model, in which the within- and between-group sampling models are both normal:

평균이 서로 다른 그룹 형태

$$\phi_j = \{\theta_j, \sigma^2\}, \;\; p(y|\phi_j) = \text{normal}(\theta_j, \sigma^2) \qquad \text{(within-group model)} \quad (8.1)$$

$$\psi = \{\mu, \tau^2\}, \;\; p(\theta_j|\psi) = \text{normal}(\mu, \tau^2) \qquad \text{(between-group model)} \quad (8.2)$$

It might help to visualize this setup as in Figure 8.3. Note that $p(\phi|\psi)$ only describes the heterogeneity across group means, and not any heterogeneity in group-specific variances. In fact, the within-group sampling variability σ^2 is assumed to be constant across groups. At the end of this chapter we will eliminate this assumption by adding a component to the model that allows for group-specific variances.

각 그룹 내의 분산은 일정하다고 가정

(파트3에서 분산 다른 경우 다룸)

Hierarchical normal model

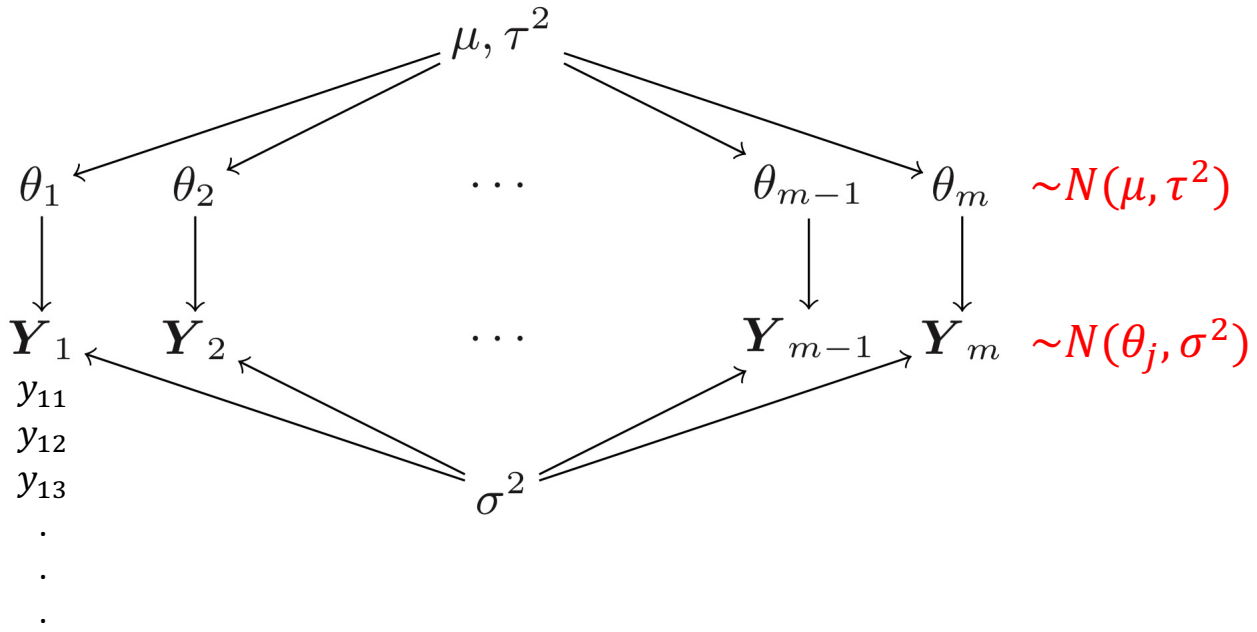


Fig. 8.3. A graphical representation of the basic hierarchical normal model.

Priors - μ, τ^2, σ^2

The fixed but unknown parameters in this model are μ, τ^2 and σ^2. For convenience we will use standard semiconjugate normal and inverse-gamma prior distributions for these parameters:

$$\begin{aligned} 1/\sigma^2 &\sim \text{ gamma } (\nu_0/2, \nu_0\sigma_0^2/2) \\ 1/\tau^2 &\sim \text{ gamma } (\eta_0/2, \eta_0\tau_0^2/2) \\ \mu &\sim \text{ normal } (\mu_0, \gamma_0^2) \end{aligned}$$

Posterior inference

Joint posterior 를 구하는 것이 목적!

by Gibbs sampling ← Full conditional distribution 구해야 함

< Joint Posterior >

$$p(\theta_1, \ldots, \theta_m, \mu, \tau^2, \sigma^2 | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$$
$$\propto p(\mu, \tau^2, \sigma^2) \times p(\theta_1, \ldots, \theta_m | \mu, \tau^2, \sigma^2) \times p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m | \theta_1, \ldots, \theta_m, \mu, \tau^2, \sigma^2)$$
$$= p(\mu)p(\tau^2)p(\sigma^2) \left\{ \prod_{j=1}^{m} p(\theta_j | \mu, \tau^2) \right\} \left\{ \prod_{j=1}^{m} \prod_{i=1}^{n_j} p(y_{i,j} | \theta_j, \sigma^2) \right\} . \qquad (8.3)$$

∵ semiconjugate
(= independent)

∵ exchangeable (= conditionally iid)
< de- Finetti's Theorm @week1 >

8.3.1 Posterior inference

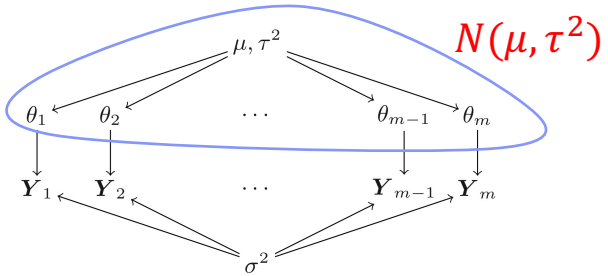
$y_i \sim \text{normal}(\mu, \sigma^2)$ σ^2 is known.	$\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}\right)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ μ is known.	$\sigma^2 \sim$ $\text{inverse gamma}(\alpha, \beta)$	$\sigma^2 \sim$ $\text{inverse gamma}\left(\alpha+\frac{n}{2}, \beta+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}\right)$

Full conditional distributions of μ and τ^2

joint posterior $\propto$ $$p(\mu)p(\tau^2)p(\sigma^2)\left\{\prod_{j=1}^{m} p(\theta_j|\mu,\tau^2)\right\}\left\{\prod_{j=1}^{m}\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)\right\}. \qquad (8.3)$$

$$p(\mu|\theta_1,\ldots,\theta_m,\tau^2,\sigma^2,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \propto p(\mu)\prod p(\theta_j|\mu,\tau^2)$$

$$p(\tau^2|\theta_1,\ldots,\theta_m,\mu,\sigma^2,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \propto p(\tau^2)\prod p(\theta_j|\mu,\tau^2).$$



$$\{\mu|\theta_1,\ldots,\theta_m,\tau^2\} \sim \text{normal}\left(\frac{m\bar{\theta}/\tau^2+\mu_0/\gamma_0^2}{m/\tau^2+1/\gamma_0^2}, [m/\tau^2+1/\gamma_0^2]^{-1}\right)$$

$$\{1/\tau^2|\theta_1,\ldots,\theta_m,\mu\} \sim \text{gamma}\left(\frac{\eta_0+m}{2}, \frac{\eta_0\tau_0^2+\sum(\theta_j-\mu)^2}{2}\right).$$

@week3
←Normal model with unknown mean

← Normal model with unknown variance

Conjugacy 이용 : normal prior + normal likelihood (var is known) ⇒ normal posterior
inverse gamma prior + normal likelihood(mean is known) ⇒ inverse gamma posterior

8.3.1 Posterior inference

Full conditional of θj

joint posterior $\propto$ $$p(\mu)p(\tau^2)p(\sigma^2)\left\{\prod_{j=1}^{m} p(\theta_j|\mu,\tau^2)\right\}\left\{\prod_{j=1}^{m}\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)\right\}. \qquad (8.3)$$

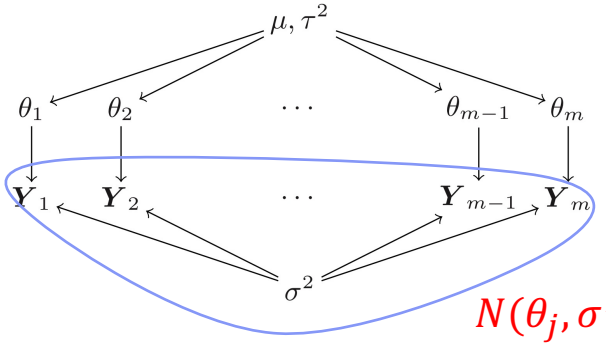
Collecting the terms in Equation 8.3 that depend on θ_j shows that the full conditional distribution of θ_j must be proportional to

$$p(\theta_j|\mu,\tau^2,\sigma^2,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \propto p(\theta_j|\mu,\tau^2)\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2).$$

$$\{\theta_j|y_{1,j},\ldots,y_{n_j,j},\sigma^2\} \sim \text{normal}\left(\frac{n_j\bar{y}_j/\sigma^2 + \mu/\tau^2}{n_j/\sigma^2+1/\tau^2}, [n_j/\sigma^2+1/\tau^2]^{-1}\right).$$

8.3.1 Posterior inference

Full conditional of σ^2



$N(\theta_j, \sigma^2)$

joint posterior $\propto$ $$p(\mu)p(\tau^2)p(\sigma^2)\left\{\prod_{j=1}^{m} p(\theta_j|\mu,\tau^2)\right\}\left\{\prod_{j=1}^{m}\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)\right\}. \qquad (8.3)$$

$$p(\sigma^2|\theta_1,\ldots,\theta_m,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \propto p(\sigma^2)\prod_{j=1}^{m}\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)$$

$$\propto (\sigma^2)^{-\nu_0/2+1} e^{-\frac{\nu_0\sigma_0^2}{2\sigma^2}} (\sigma^2)^{-\sum n_j/2} e^{-\frac{\sum\sum(y_{i,j}-\theta_j)^2}{2\sigma^2}}.$$

$$\{1/\sigma^2|\boldsymbol{\theta},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n\} \sim \text{gamma}\left(\frac{1}{2}[\nu_0+\sum_{j=1}^{m} n_j], \frac{1}{2}[\nu_0\sigma_0^2+\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2]\right).$$

8.3.1 Posterior inference

Full conditionals 정리

$$\{\mu|\theta_1,\ldots,\theta_m,\tau^2\} \sim \text{normal}\left(\frac{m\bar{\theta}/\tau^2+\mu_0/\gamma_0^2}{m/\tau^2+1/\gamma_0^2}, [m/\tau^2+1/\gamma_0^2]^{-1}\right)$$

$$\{1/\tau^2|\theta_1,\ldots,\theta_m,\mu\} \sim \text{gamma}\left(\frac{\eta_0+m}{2}, \frac{\eta_0\tau_0^2+\sum(\theta_j-\mu)^2}{2}\right).$$

$$\{\theta_j|y_{1,j},\ldots,y_{n_j,j},\sigma^2\} \sim \text{normal}(\frac{n_j\bar{y}_j/\sigma^2+\mu/\tau^2}{n_j/\sigma^2+1/\tau^2}, [n_j/\sigma^2+1/\tau^2]^{-1}).$$

$$\{1/\sigma^2|\boldsymbol{\theta},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n\} \sim \text{gamma}(\frac{1}{2}[\nu_0+\sum_{j=1}^{m}n_j], \frac{1}{2}[\nu_0\sigma_0^2+\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2]).$$

8.4

Example: Math scores in U.S. public schools

8.4 Example: Math scores in U.S. public schools

Data

- 10th grade children from 100 different large urban public high schools, all having a 10th grade enrollment of 400 or greater.
- Data from these schools are shown in Figure 8.4, with scores from students within the same school plotted along a common vertical bar.

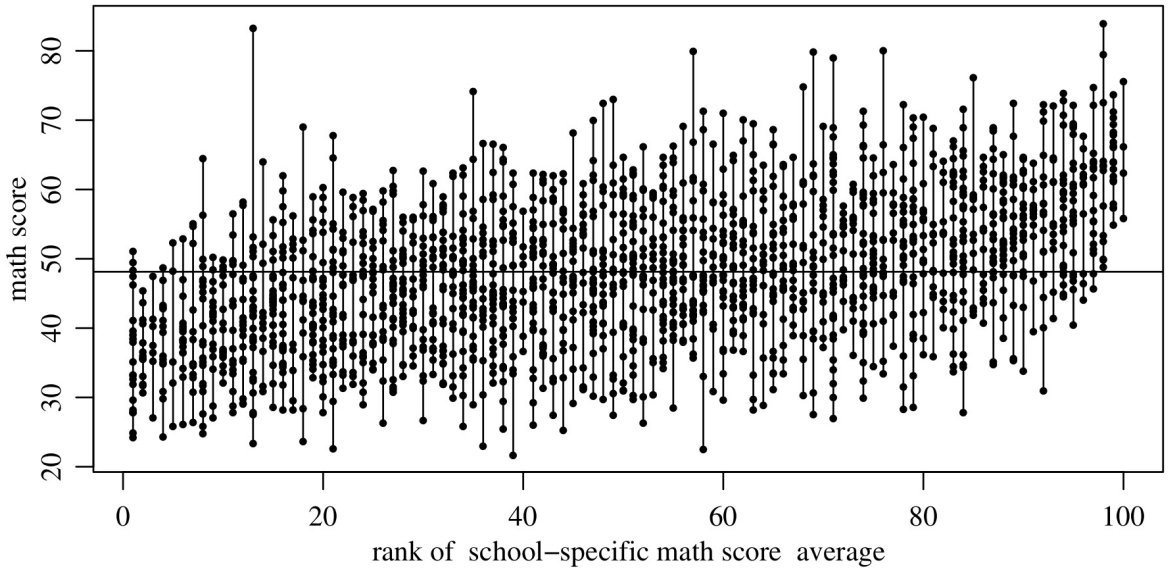


Fig. 8.4. A graphical representation of the ELS data.

8.4 Example: Math scores in U.S. public schools

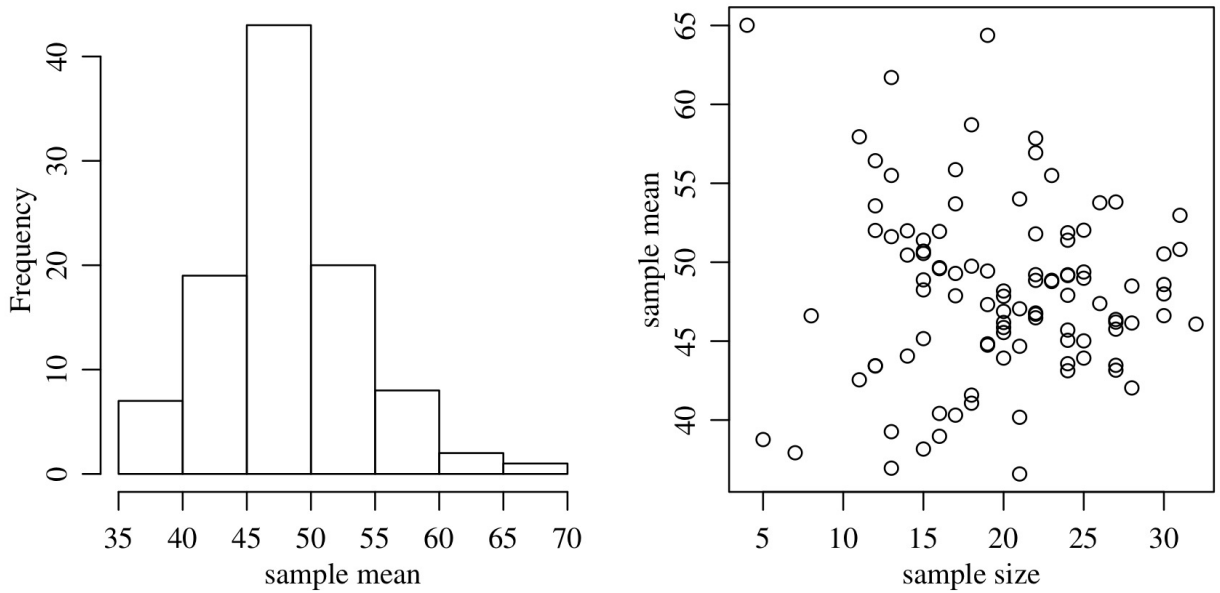


Fig. 8.5. Empirical distribution of sample means, and the relationship between sample mean and sample size.

- The range of average scores is quite large, with the lowest average being 36.6 and the highest 65.0.
- Very extreme sample averages tend to be associated with schools with small sample sizes.

sample variance = $\mathrm{Var}[\bar{Y}_j|\sigma_j^2] = \sigma^2/n_j$

8.4.1 Prior distributions and posterior approximation

Gibbs sampling 개요

The prior parameters we need to specify are

(ν_0, σ_0^2) for $p(\sigma^2)$,
(η_0, τ_0^2) for $p(\tau_0^2)$ and
(μ_0, γ_0^2) for $p(\mu)$.

Posterior approximation proceeds by iterative sampling of each unknown quantity from its full conditional distribution. Given a current state of the unknowns $\{\theta_1^{(s)}, \ldots, \theta_m^{(s)}, \mu^{(s)}, \tau^{2(s)}, \sigma^{2(s)}\}$, a new state is generated as follows:

1. sample $\mu^{(s+1)} \sim p(\mu|\theta_1^{(s)}, \ldots, \theta_m^{(s)}, \tau^{2(s)})$;
2. sample $\tau^{2(s+1)} \sim p(\tau^2|\theta_1^{(s)}, \ldots, \theta_m^{(s)}, \mu^{(s+1)})$;
3. sample $\sigma^{2(s+1)} \sim p(\sigma^2|\theta_1^{(s)}, \ldots, \theta_m^{(s)}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$;
4. for each $j \in \{1, \ldots, m\}$, sample $\theta_j^{(s+1)} \sim p(\theta_j|\mu^{(s+1)}, \tau^{2(s+1)}, \sigma^{2(s+1)}, \boldsymbol{y}_j)$.

8.4.1 Prior distributions and posterior approximation

Gibbs sampling code_initial settings

```
### weakly informative priors
nu0<-1  ; s20<-100
eta0<-1 ; t20<-100
mu0<-50 ; g20<-25
###

### starting values
m<-length(unique(Y[,1]))
n<-sv<-ybar<-rep(NA,m)
for(j in 1:m)
{
  ybar[j]<-mean(Y[Y[,1]==j,2])
  sv[j]<-var(Y[Y[,1]==j,2])
  n[j]<-sum(Y[,1]==j)
}
theta<-ybar ; sigma2<-mean(sv)
mu<-mean(theta) ; tau2<-var(theta)
###

### setup MCMC
set.seed(1)
S<-5000
THETA<-matrix( nrow=S,ncol=m)
SMT<-matrix( nrow=S,ncol=3)
###
```

The prior parameters we need to specify are

(ν_0, σ_0^2) for $p(\sigma^2)$,
(η_0, τ_0^2) for $p(\tau_0^2)$ and
(μ_0, γ_0^2) for $p(\mu)$.

8.4.1 Prior distributions and posterior approximation

Gibbs sampling code _MCMC algorithm

$$\{\theta_j|y_{1,j},\ldots,y_{n_j,j},\sigma^2\} \sim \text{normal}(\frac{n_j\bar{y}_j/\sigma^2 + \mu/\tau^2}{n_j/\sigma^2+1/\tau^2}, [n_j/\sigma^2+1/\tau^2]^{-1}).$$

$$\{1/\sigma^2|\boldsymbol{\theta},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n\} \sim \text{gamma}(\frac{1}{2}[\nu_0+\sum_{j=1}^{m} n_j], \frac{1}{2}[\nu_0\sigma_0^2+\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2]).$$

$$\{\mu|\theta_1,\ldots,\theta_m,\tau^2\} \sim \text{normal}\left(\frac{m\bar{\theta}/\tau^2+\mu_0/\gamma_0^2}{m/\tau^2+1/\gamma_0^2}, [m/\tau^2+1/\gamma_0^2]^{-1}\right)$$

$$\{1/\tau^2|\theta_1,\ldots,\theta_m,\mu\} \sim \text{gamma}\left(\frac{\eta_0+m}{2}, \frac{\eta_0\tau_0^2+\sum(\theta_j-\mu)^2}{2}\right).$$

```
### MCMC algorithm
for(s in 1:S)
{

  # sample new values of the thetas
  for(j in 1:m)
  {
    vtheta<-1/(n[j]/sigma2+1/tau2)
    etheta<-vtheta*(ybar[j]*n[j]/sigma2+mu/tau2)
    theta[j]<-rnorm(1,etheta,sqrt(vtheta))
  }

  #sample new value of sigma2
  nun<-nu0+sum(n)
  ss<-nu0*s20
  for(j in 1:m){ss<-ss+sum((Y[Y[,1]==j,2]-theta[j])^2)}
  sigma2<-1/rgamma(1,nun/2,ss/2)

  #sample a new value of mu
  vmu<- 1/(m/tau2+1/g20)
  emu<- vmu*(m*mean(theta)/tau2 + mu0/g20)
  mu<-rnorm(1,emu,sqrt(vmu))

  # sample a new value of tau2
  etam<-eta0+m
  ss<- eta0*t20 + sum( (theta-mu)^2 )
  tau2<-1/rgamma(1,etam/2,ss/2)

  #store results
  THETA[s,]<-theta
  SMT[s,]<-c(sigma2,mu,tau2)

}
###
```

8.4.2 Posterior summaries and shrinkage

Result: Posterior densities_ μ, τ^2, σ^2

Figure 8.7 shows Monte Carlo approximations to the posterior densities of $\{\mu, \sigma^2, \tau^2\}$. The posterior means of μ, σ and τ are 48.12, 9.21 and 4.97 respec-

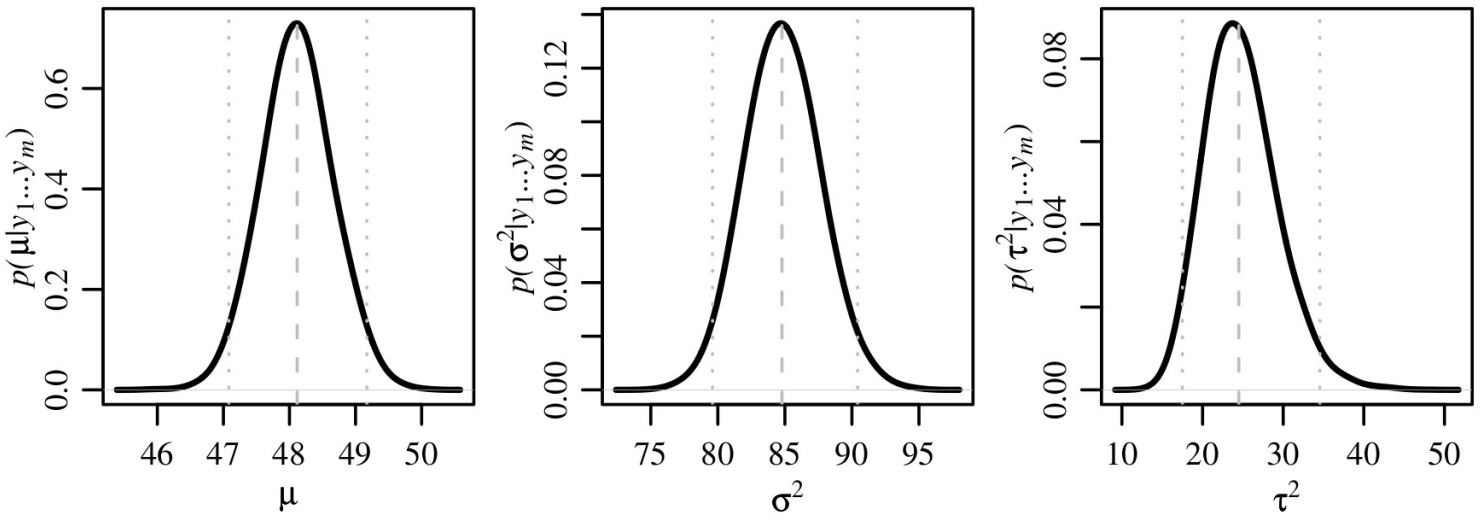


Fig. 8.7. Marginal posterior distributions, with 2.5%, 50% and 97.5% quantiles given by vertical lines.

8.4.2 Posterior summaries and shrinkage

Posterior_θj

$$\{\theta_j|y_{1,j},\ldots,y_{n_j,j},\sigma^2\} \sim \text{normal}(\frac{n_j\bar{y}_j/\sigma^2 + \mu/\tau^2}{n_j/\sigma^2+1/\tau^2}, [n_j/\sigma^2+1/\tau^2]^{-1}).$$

Posterior mean = $\mathrm{E}[\theta_j|\boldsymbol{y}_j,\mu,\tau,\sigma] = \frac{\bar{y}_j n_j/\sigma^2+\mu/\tau^2}{n_j/\sigma^2+1/\tau^2}$ · (Sample mean과 prior mean의 precision에 대한 가중평균 형태)

As a result, the expected value of θ_j is pulled a bit from $\bar{y}_j$ towards μ by an amount depending on n_j.
This effect is called **shrinkage**.

8.4.2 Posterior summaries and shrinkage

Shrinkage

high values of $\bar{y}_j$ correspond to slightly less high values of $\hat{\theta}_j$
low values of $\bar{y}_j$ correspond to slightly less low values of $\hat{\theta}_j$.

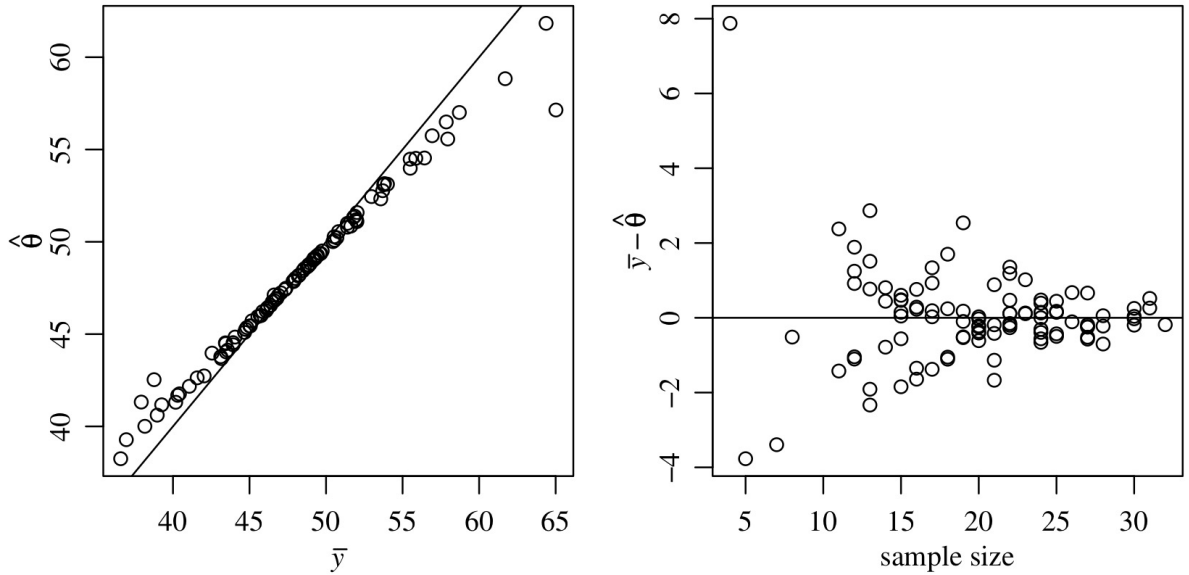


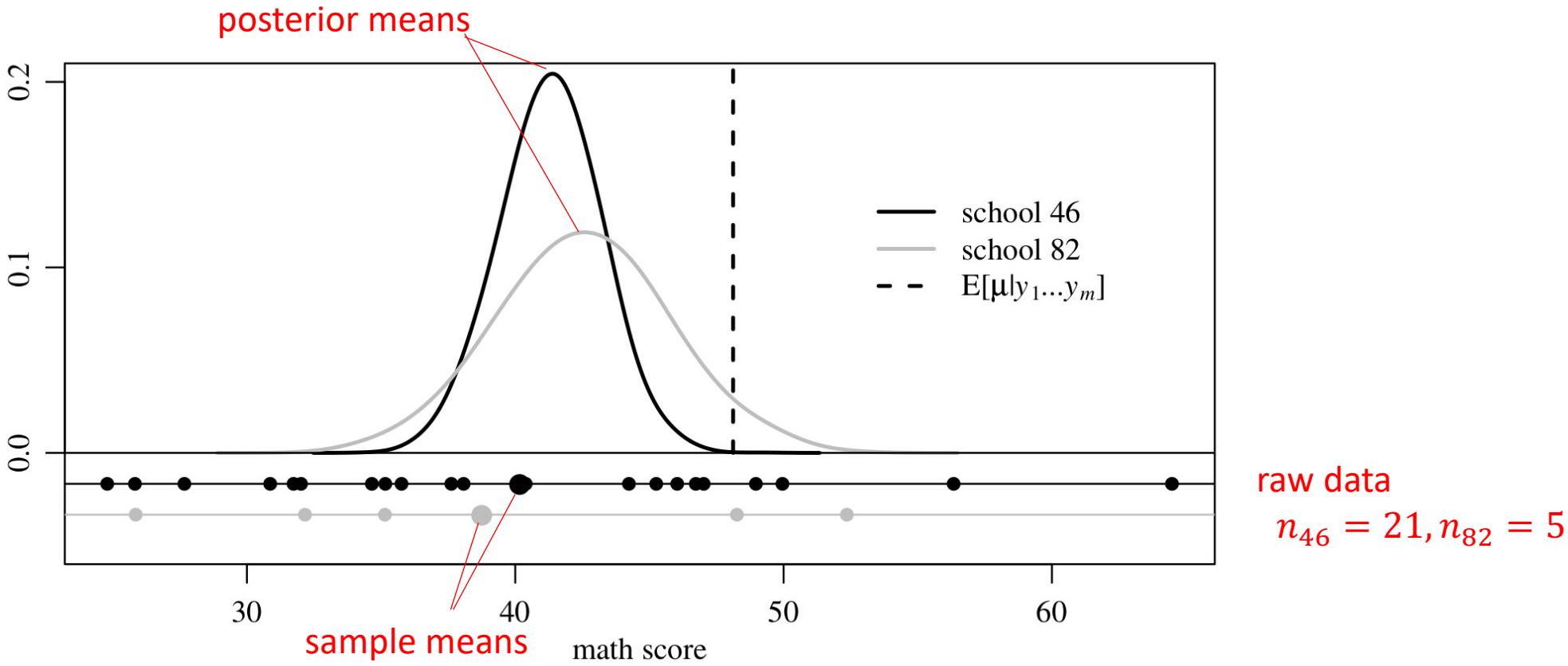
Fig. 8.8. Shrinkage as a function of sample size.

* One of the motivations behind hierarchical modeling is that information can be shared across groups. *

- ✓ Groups with low sample sizes get shrunk the most, whereas groups with large sample sizes hardly get shrunk at all.
- ✓ The **larger the sample size** for a group, the more information we have for that group and the **less information we need to "borrow"** from the rest of the population.

8.4.2 Posterior summaries and shrinkage

Shrinkage 예시



raw data

$n_{46} = 21, n_{82} = 5$

Fig. 8.9. Data and posterior distributions for two schools.

Part3
Hierarchical modeling of means and variance

FCB 8.5

- σ_j^2 : variance for group j

- Sampling model

  $Y_{1,j}, \cdots, Y_{n_j,j} \sim$ i.i.d normal (θ_j, σ_j^2)

- Full conditional distribution for each θ_j

$$\{\theta_j | \, y_{1,j}, \cdots, y_{n_j,j}, \, \sigma_j^2\} \sim \text{normal}\left(\frac{n_j \bar{y}_j / \sigma_j^2 + 1/\tau^2}{n_j/\sigma_j^2 + 1/\tau^2} \, , \, [n_j/\sigma_j^2 + 1/\tau^2]^{-1} \right)$$

Estimation of σ_j^2

- $\sigma_1^2, \cdots, \sigma_m^2 \sim$ i.i.d gamma($\nu_0/2, \nu_0\sigma_0^2/2$) *m = 그룹 수

↓ chapter6

- Full conditional distribution of σ_j^2

$$\left\{\frac{1}{\sigma_j^2} \mid y_{1,j}, \cdots, y_{n_j,j}, \theta_j\right\} \sim \text{gamma}\left(\frac{[\nu_0 + n_j]}{2}, \frac{[\nu_0\sigma_0^2 + (y_{i,j} - \theta_j)^2]}{2}\right)$$

chapter 6.3 sampling from the conditional distributions

Given θ_j, $\{y_{1,j}, \cdots, y_{n_j,j}\}$

$p\{\frac{1}{\sigma_j^2} | \ y_{1,j}, \cdots, y_{n_j,j}, \theta_j\}$

$\propto p(y_{1,j}, \cdots, y_{n_j,j}, \ \theta_j, \frac{1}{\sigma_j^2}) = p(y_{1,j}, \cdots, y_{n_j,j} | \ \theta_j, \frac{1}{\sigma_j^2}) p(\theta_j \ | \frac{1}{\sigma_j^2}) p(\frac{1}{\sigma_j^2})$

$\propto p(y_{1,j}, \cdots, y_{n_j,j} | \ \theta_j, \frac{1}{\sigma_j^2}) p(\frac{1}{\sigma_j^2})$

$\propto ((\frac{1}{\sigma_j^2})^{n_j/2} \exp\{-\frac{1}{\sigma_j^2} \sum_{i=1}^{n_j} (y_{i,j} - \ \theta_j)^2/2\}) \times ((\frac{1}{\sigma_j^2})^{\nu_0/2-1} \ \exp\{-\frac{1}{\sigma_j^2} \nu_0 \sigma_0^2/2\})$

$= (\frac{1}{\sigma_j^2})^{\nu_0 + n_j/2 - 1} \times \exp\{-\frac{1}{\sigma_j^2} \times [\ \nu_0 \sigma_0^2 + \ \sum_{i=1}^{n_j} (y_{i,j} - \ \theta_j)^2]/2\}$

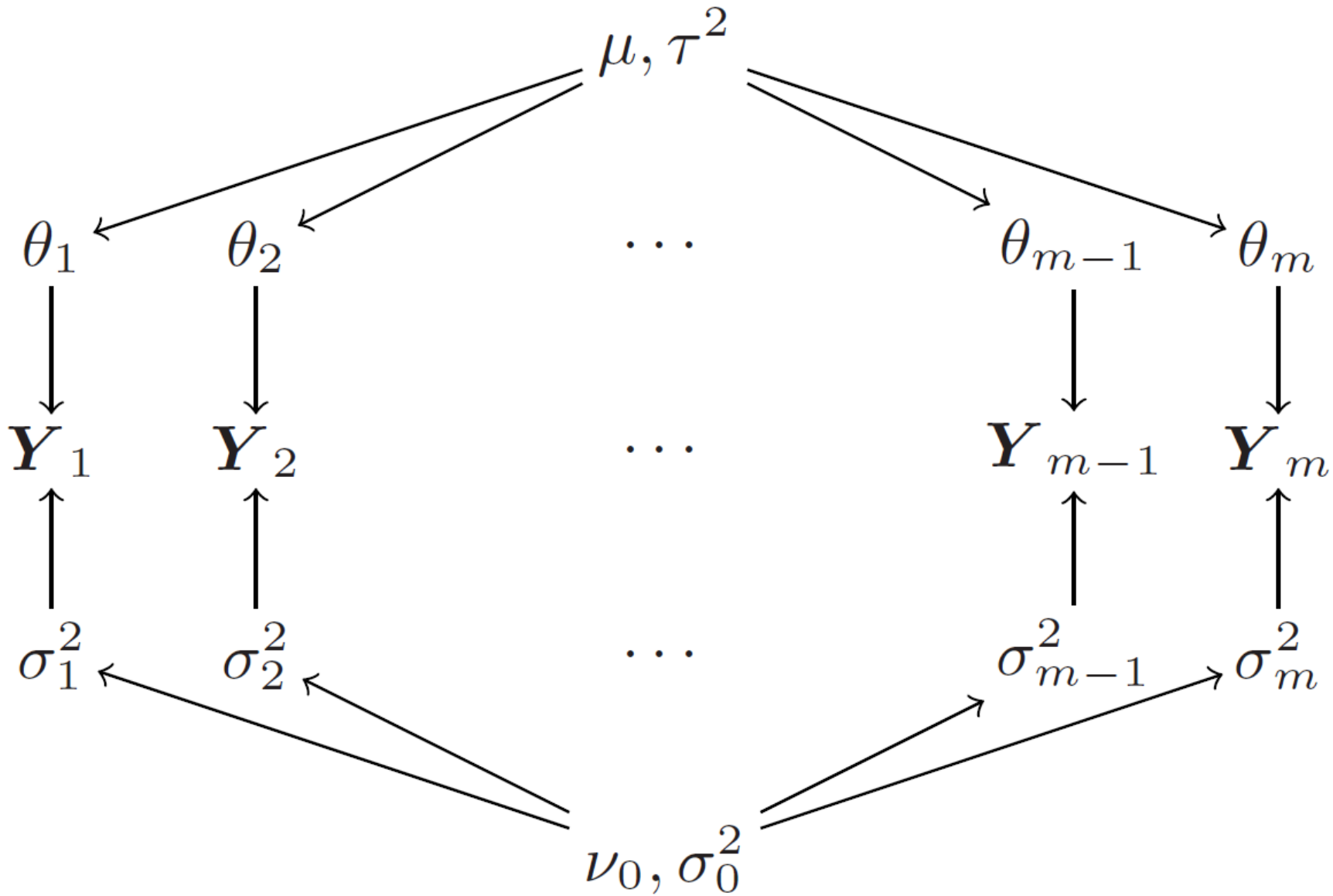
→ gamma density

$\therefore \{\sigma_j^2 \mid y_{1,j}, \cdots, y_{n_j,j}, \theta_j\} \sim IG(\nu_0 + n_j/2, \ [\nu_0 \sigma_0^2 + \sum_{i=1}^{n_j} (y_{i,j} - \ \theta_j)^2]/2)$

About ν_0, σ_0^2

- If ν_0 , σ_0^2 are fixed at particular value in advance
  → $\sigma_1^2, \cdots, \sigma_m^2$ ~ i.i.d gamma($\nu_0/2$, $\nu_0\sigma_0^2/2$) is a prior distribution.
  → $p(\sigma_m^2 | \sigma_1^2, \cdots, \sigma_{m-1}^2) = p(\sigma_m^2)$. $\sigma_1^2, \cdots, \sigma_{m-1}^2$ useless.

- If ν_0 , σ_0^2 are parameters to be estimated
  → $\sigma_1^2, \cdots, \sigma_m^2$ ~ i.i.d gamma($\nu_0/2$, $\nu_0\sigma_0^2/2$) is a sampling model for across-group heterogeneity in population variance.
  → small sample size,
  $\sigma_1^2, \cdots, \sigma_{m-1}^2$ were tightly concentrated around particular value, improve estimation of σ_m^2


μ, τ^2
θ_1
θ_2
$\cdots$
θ_{m-1}
θ_m
$\boldsymbol{Y}_1$
$\boldsymbol{Y}_2$
$\cdots$
$\boldsymbol{Y}_{m-1}$
$\boldsymbol{Y}_m$
σ_1^2
σ_2^2
$\cdots$
σ_{m-1}^2
σ_m^2
ν_0, σ_0^2

Unknown parameter

- $\{(\theta_1,\sigma_1^2),\cdots,(\theta_m,\sigma_m^2)\}$: within-group sampling distribution

$$\{\theta_j|\ y_{1,j},\cdots,y_{n_j,j},\ \sigma_j^2\} \sim \text{normal}\left(\frac{n_j\bar{y}_j/\sigma_j^2+1/\tau^2}{n_j/\sigma_j^2+1/\tau^2},\ [n_j/\sigma_j^2+1/\tau^2]^{-1}\right)$$

$$\{\frac{1}{\sigma_j^2}|\ y_{1,j},\cdots,y_{n_j,j},\theta_j\} \sim \text{gamma}\left(\frac{[\nu_0+n_j]}{2},\ \frac{[\nu_0\sigma_0^2+\ (y_{i,j}-\ \theta_j)\ ^2]}{2}\right)$$

- $\{\mu\ ,\tau^2\}$: across-group heterogeneity in means

$$\{\mu\ |\theta_1,\cdots,\theta_2,\tau^2\} \sim \text{normal}\left(\frac{m\bar{\theta}/\tau^2+\mu_0/{\gamma_0}^2}{m/\tau^2+1/{\gamma_0}^2},\ [m/\tau^2+1/{\gamma_0}^2]^{-1}\right)$$

$$\{1/\tau^2|\ \theta_1,\cdots,\theta_2,\mu\}\sim\text{gamma}\left(\frac{\eta_0+m}{2},\ \frac{\eta_0\tau_0^2+\sum(\theta_j-\mu)^2}{2}\right)$$

- $\{\nu_0,\sigma_0^2\}$: across-group heterogeneity in variance - ??

Prior distributions and Full conditional distribution for ν_0, σ_0^2

- σ_0^2

  conjugate class of prior densities : gamma

  → if p(σ_0^2)~gamma(a,b)

  then p(σ_0^2| σ_1^2,⋯, σ_m^2,ν_0)=gamma(a+$\frac{1}{2}$mν_0, b+ $\frac{1}{2}\sum_{j=1}^{m}\frac{1}{\sigma_j^2}$)

  *for small a and b,

  conditional mean of σ_0^2 ≒ harmonic mean of σ_1^2,⋯, σ_m^2

Prior distributions and Full conditional distribution for ν_0, σ_0^2

- ν_0

  ∄simple conjugate prior

  → if restrict ν_0 to whole number

  then easily sample from full conditional distribution

  example) prior on ν_0 : geometric distribution on {1,2,3,…}

  → $p(\nu_0) \propto e^{-\alpha\nu_0}$

  → $p(\nu_0|\sigma_0^2,\sigma_1^2,\cdots,\sigma_m^2)$

  $\propto p(\nu_0) \times p(\sigma_1^2,\cdots,\sigma_m^2|\nu_0,\sigma_0^2)$

  $\propto \left(\frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right)^m \left(\prod_{j=1}^m \frac{1}{\sigma_j^2}\right)^{\nu_0/2-1} \times \exp\{-\nu_0(\alpha+\frac{1}{2}\sigma_0^2\sum(1/\sigma_j^2))\}$

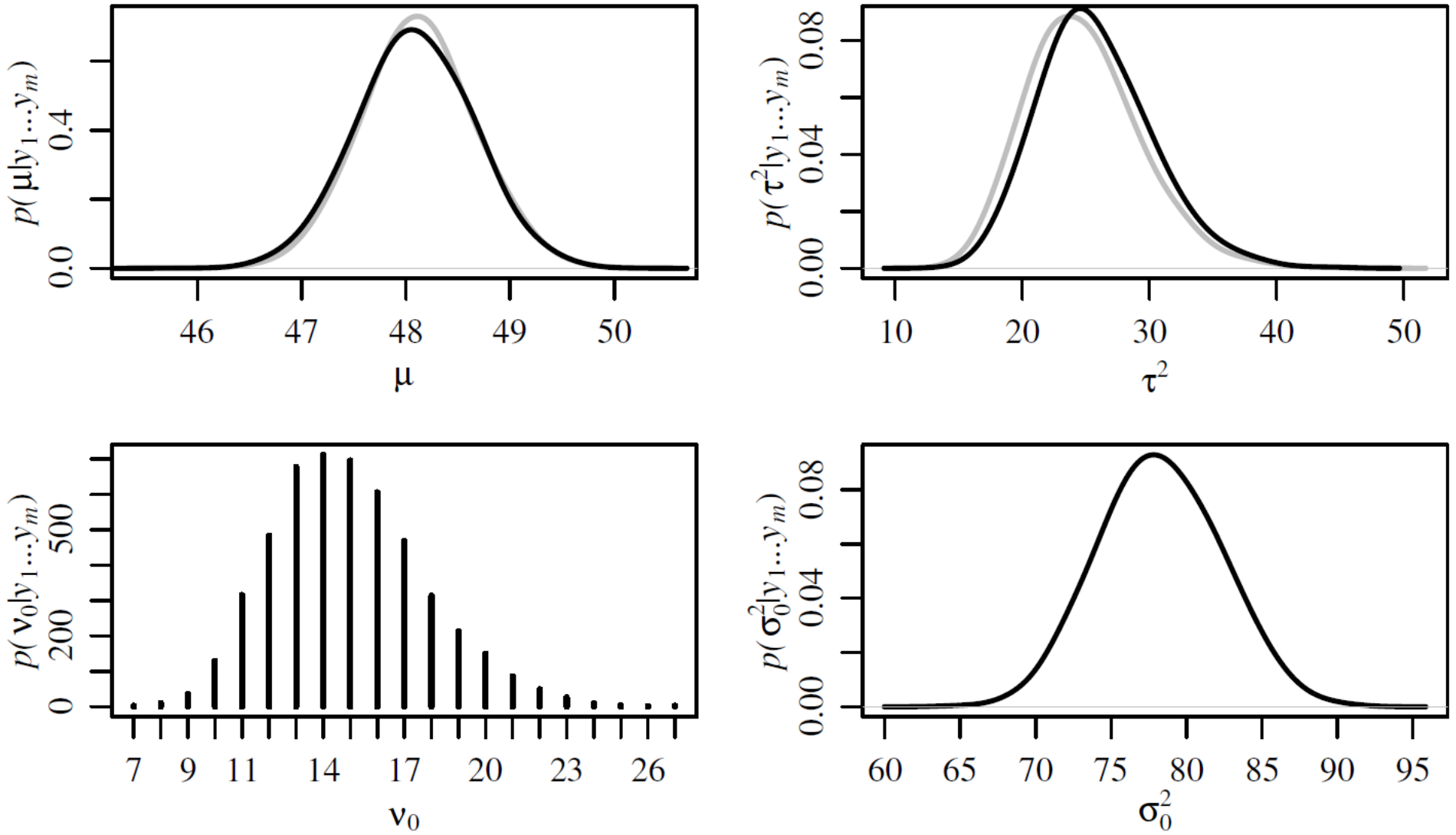
Gibbs sampling code

```
#### Hierarchicl model for the mean and variance

## weakly informative priors
nu0<-1  ; s20<-100
eta0<-1 ; t20<-100
mu0<-50 ; g20<-25
a0<-1 ; b0<-1/100 ; wnu0<-1

## starting values
m<-length(Y)
n<-sv<-ybar<-rep(NA,m)
for(j in 1:m)
{
  ybar[j]<-mean(Y[[j]])
  sv[j]<-var(Y[[j]])
  n[j]<-length(Y[[j]])
}
theta<-ybar
sigma2<-sv
mu<-mean(theta)
tau2<-var(theta)
s20<-1/mean(1/sv)
nu0<-10


## setup MCMC
set.seed(1)
S<-5000
SIGMA2<-THETA<-matrix( nrow=S,ncol=m)
MTSN<-matrix( nrow=S,ncol=4)
s2.pp<-NULL
nu0s<-1:5000
```

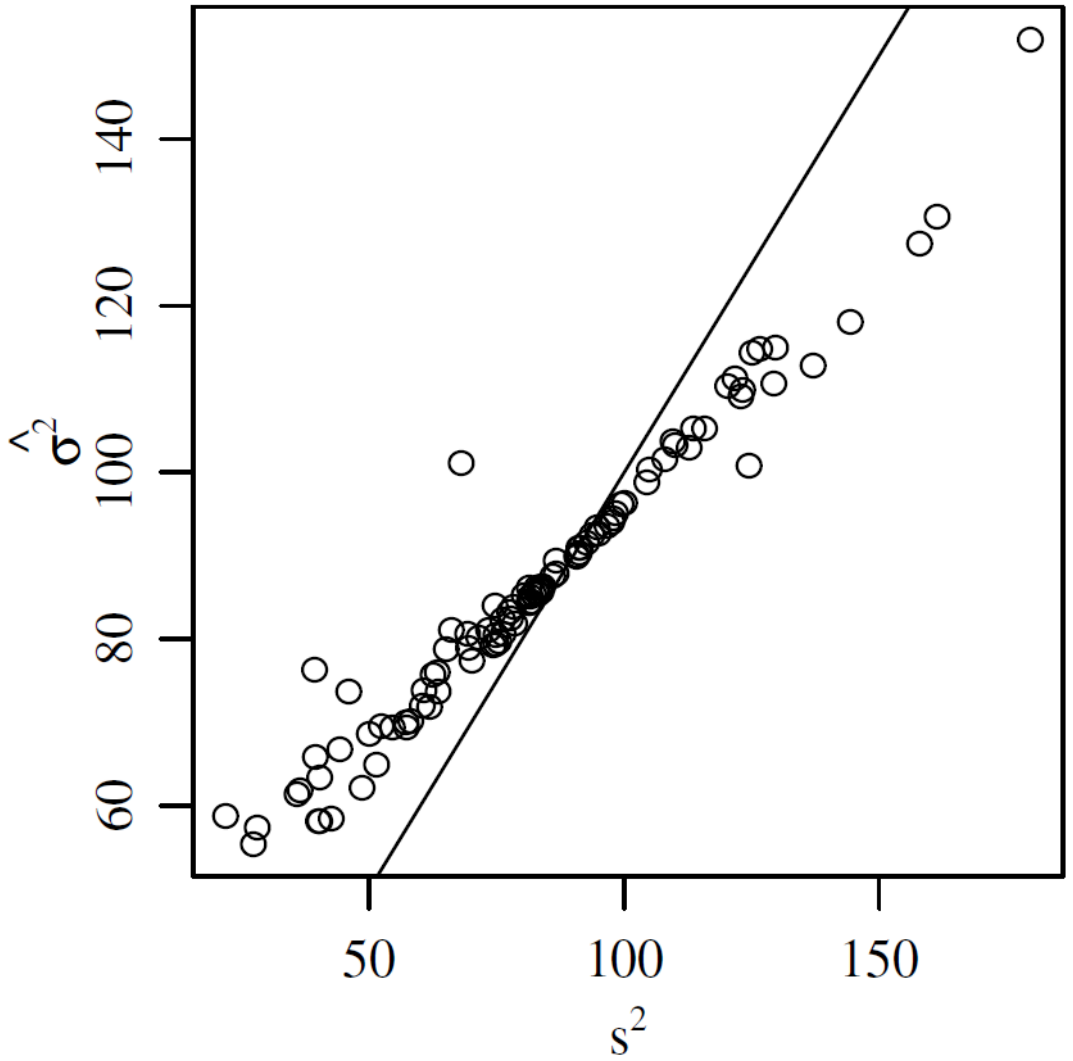
```
## MCMC algorithm
for(s in 1:S)
{
  # sample new values of the thetas
  for(j in 1:m)
  {
    vtheta<-1/(n[j]/sigma2[j]+1/tau2)
    etheta<-vtheta*(ybar[j]*n[j]/sigma2[j]+mu/tau2)
    theta[j]<-rnorm(1,etheta,sqrt(vtheta))
   }

  #sample new value the sigma2s
  for(j in 1:m)
  {
  nun<-nu0+n[j]
  ss<-nu0*s20+ sum((Y[[j]]-theta[j])^2)
  sigma2[j]<-1/rgamma(1,nun/2,ss/2)
  }

  #sample new s20
  s20<-rgamma(1,a0+m*nu0/2,b0+nu0*sum(1/sigma2)/2)

  lpnu0<- .5*nu0s*m*log(s20*nu0s/2)-m*lgamma(nu0s/2)+(nu0s/2-1)*sum(log(1/sigma2)) -
          nu0s*s20*sum(1/sigma2)/2   - wnu0*nu0s
  nu0<-sample(nu0s,1,prob=exp( lpnu0-max(lpnu0)) )

  #sample a new value of mu
  vmu<- 1/(m/tau2+1/g20)
  emu<- vmu*(m*mean(theta)/tau2 + mu0/g20)
  mu<-rnorm(1,emu,sqrt(vmu))

  # sample a new value of tau2
  etam<-eta0+m
  ss<- eta0*t20 + sum( (theta-mu)^2 )
  tau2<-1/rgamma(1,etam/2,ss/2)

  #store results
  THETA[s,]<-theta
  SIGMA2[s,]<-sigma2
  MTSN[s,]<-c(mu,tau2,s20,nu0)
  s2.pp<-c(s2.pp,1/rgamma(1,nu0/2,nu0*s20/2))
  if(s %%25 ==0) { hist(1/sigma2,prob=T) ;
       x<-seq(0,20,100); lines(x,dgamma(x,nu0/2,nu0*s20/2)) }

}
```

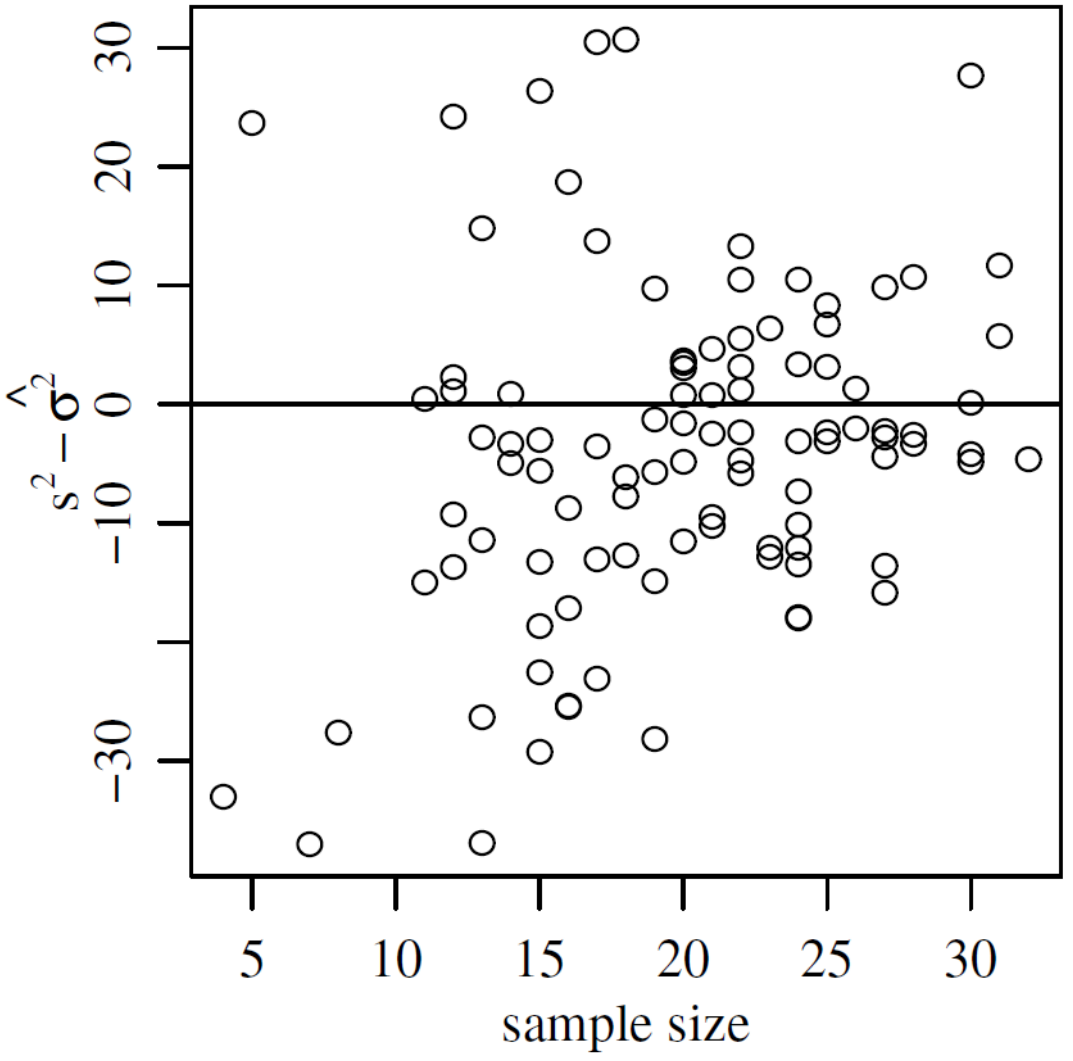
Analysis of math score data

- α=1, {a=1, b=100}, Gibbs sampler for 5000 iterations



shrinkage


$\hat{\sigma}^2$
60
80
100
120
140
50
100
150
s^2


$s^2 - \hat{\sigma}^2$
−30
−10
0
10
20
30
5
10
15
20
25
30
sample size

part2 8.4 Gibbs sampling code 참고

HW

week5hw.R의 코드를 직접 R로 실행하며 과제 수행해주시면 됩니다!

<코드 흐름>

0-1. 데이터와 초기 세팅

part2 8.4장에서 다룬 math score example에 대한 데이터입니다.

2. MCMC algorithm

✓ "###채우기###" 부분 채워넣기

✓ 손으로 수식 작성해보기
코드가 뭘 나타내는건지 직접 수식으로 작성해보고 conjugacy를 이용하여 해당 형태가 어떻게 도출되었는지 분석해보기
(prior mean, sample mean, precision...)

Hint: PPT Week5 part2 <Full conditionals 정리> 슬라이드 참고

```
#------------------------------------------------------------------------#
##### 2. MCMC algorithm #####
#------------------------------------------------------------------------#
for(s in 1:S)
{

  # sample new values of the thetas (within-group model의 서로 다른 평균들)
  for(j in 1:m)
  {
    vtheta <- ### 채우기 ###
    etheta <- vtheta*(ybar[j]*n[j]/sigma2+mu/tau2)              #손으로 수식 작성해보기
    theta[j] <- rnorm(1,etheta,sqrt(vtheta))
  }

  #sample new value of sigma2 (within-group model의 분산 - 모두 동일한 경우)
  nun <- ### 채우기 ###
  ss <- nu0*s20;for(j in 1:m){ss<-ss+sum((Y[[j]]-theta[j])^2)}   #손으로 수식 작성해보기
  sigma2 <- 1/rgamma(1,nun/2,ss/2)

  #sample a new value of mu (between-group model의 평균)
  vmu <- 1/(m/tau2+1/g20)
  emu <- vmu*(m*mean(theta)/tau2 + mu0/g20)
  mu <- ### 채우기 ###

  # sample a new value of tau2 (between-group model의 분산)
  etam <- eta0+m
  ss <- eta0*t20 + sum( (theta-mu)^2 )
  tau2 <- ### 채우기 ###
```

3. 교재에 있는 plot 직접 그려보기

코드를 돌리면서 Figure 8.7: Posterior distributions, Figure 8.8: Shrinkage 두 plot을 직접 그려보시면 됩니다!