STAT3016/4116/7016: Introduction to Bayesian Data Analysis

RSFAS, College of Business and Economics, ANU

Group Comparisons and Hierarchical Modelling

Introduction

cannot still use multivariate normal.

(frequentists' approach: pairs t-test for the same group; difference-mean test for differ groups.)

1. Two group comparison
2. Multigroup comparison - hierarchical model
3. Across group heterogeneity in variances

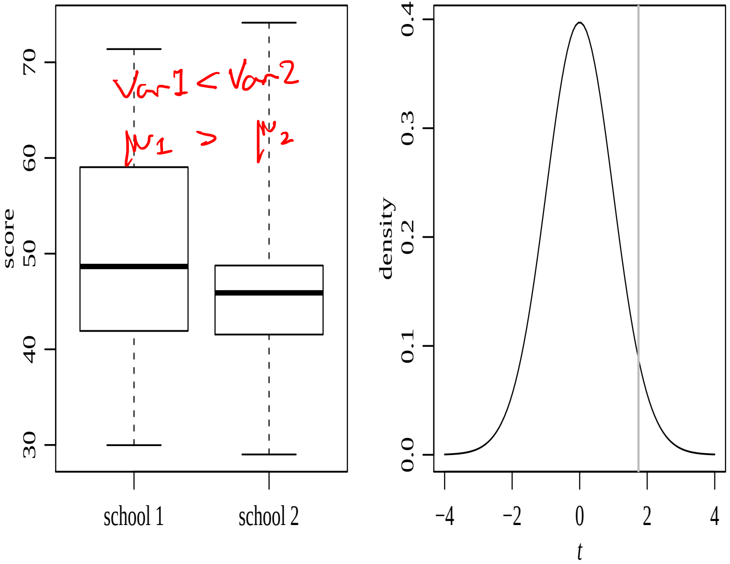
assumption: within group variance are the same. (for hierarchical model)

within group variances are different.

Comparing two groups - example

31 students   28 students

twelch →
obs. statistics



Comparing two groups - example

Frequentist approach

Welch Two sample t-test: test if the means of two groups are the same.

```
> t.test(y1,y2)

Welch Two Sample t-test

data:  y1 and y2
t = 1.7612, df = 56.288, p-value = 0.08363
alternative hypothesis: true difference in means
    is not equal to 0
95 percent confidence interval:
 -0.640171  9.965839
sample estimates:
mean of x mean of y
 50.81355  46.15071
```

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

do not reject H_0.

So $\mu_1 = \mu_2 \Rightarrow$

It's hard to say since our conclusion really depends on the significance level alpha.

Is the t-statistic large compared to the sampling variability??

Comparing two groups - Bayesian approach

Allow for information to be shared across the groups.

common mean

$$Y_{i,1} = \mu + \delta + \epsilon_{i,1} \tag{1}$$

deviation

$$Y_{i,2} = \mu - \delta + \epsilon_{i,2} \tag{2}$$

$$\epsilon_{i,1} \overset{\text{iid}}{\sim} \text{normal}(0, \sigma^2) \tag{3}$$

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

$\theta_1 = \mu + \delta$; $\theta_2 = \mu - \delta$;
$\delta = (\theta_1 - \theta_2)/2$; $\mu = (\theta_1 + \theta_2)/2$

Basic structure of 2 group comparison model

Comparing two groups - Bayesian approach

for simplicity, they're independent

Conjugate priors

$$p(\mu, \delta, \sigma^2) = p(\mu) \times p(\delta) \times p(\sigma^2)$$

$$\mu \sim \text{normal}(\mu_0, \gamma_0^2)$$

rely on Gibbs sampler to do posterior draws.

$$\delta \sim \text{normal}(\delta_0, \tau_0^2)$$

→ measures between group variation

$$\sigma^2 \sim \text{inverse-gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$

→ within group variation

Comparing two groups - Bayesian approach

plug in μ_n for μ,

σ_n^2 for σ

$\gamma_0 = \tau_0^2 = \infty$

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

Full conditional distributions

$\{\mu | \mathbf{y}_1, \mathbf{y}_2, \delta, \sigma^2\} \sim \text{normal}(\mu_n, \gamma_n^2)$ where

$\mu_n = \gamma_n^2 \times \left[\mu_0/\gamma_0^2 + \sum_{i=1}^{n_1}(y_{i,1} - \delta)/\sigma^2 + \sum_{i=1}^{n_2}(y_{i,2} + \delta)/\sigma^2\right]$

$\gamma_n^2 = \left[1/\gamma_0^2 + (n_1 + n_2)/\sigma^2\right]^{-1}$

$\{\delta | \mathbf{y}_1, \mathbf{y}_2, \mu, \sigma^2\} \sim \text{normal}(\delta_n, \tau_n^2)$ where

$\delta_n = \tau_n^2 \times \left[\delta_0/\tau_0^2 + \sum_{i=1}^{n_1}(y_{i,1} - \mu)/\sigma^2 - \sum_{i=1}^{n_2}(y_{i,2} - \mu)/\sigma^2\right]$

$\tau_n^2 = \left[1/\tau_0^2 + (n_1 + n_2)/\sigma^2\right]^{-1}$

$\{\sigma^2 | \mathbf{y}_1, \mathbf{y}_2, \mu, \delta\} \sim \text{inverse-gamma}(\nu_n/2, \nu_n\sigma_n^2/2)$ where

$\nu_n = \nu_0 + n_1 + n_2$

$\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + \sum_{i=1}^{n_1}(y_{i,1} - [\mu + \delta])^2 - \sum_{i=1}^{n_2}(y_{i,2} - [\mu - \delta])^2$

when $\nu_0 = 0$

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

Comparing two groups - Bayesian approach

Analysis of the math score data

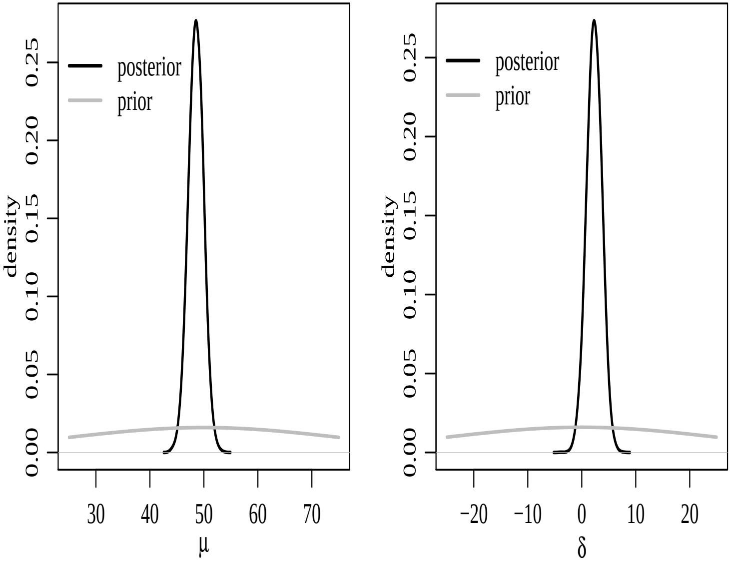
$\mu_0 = 50$; $\sigma_0^2 = 10^2 = 100$; $\gamma_0^2 = 25^2 = 625$; $\nu_0 = 1$
$\delta_0 = 0$; $\tau_0^2 = 25^2 = 625$.

```
##### prior parameters
mu0<-50 ; g02<-625
del0<-0 ; t02<-625
s20<-100; nu0<-1
#####

##### starting values
mu<- ( mean(y1) + mean(y2) )/2
del<- ( mean(y1) - mean(y2) )/2
#####
```

Comparing two groups - Bayesian approach

```
##### Gibbs sampler
MU<-DEL<-S2<-NULL
Y12<-NULL
set.seed(1)
for(s in 1:5000)
{
 ##update s2
  s2<-1/rgamma(1,(nu0+n1+n2)/2,
        (nu0*s20+sum((y1-mu-del)^2)+sum((y2-mu+del)^2) )/2)
  ##update mu
  var.mu<-  1/(1/g02+ (n1+n2)/s2 )
  mean.mu<- var.mu*( mu0/g02 + sum(y1-del)/s2 + sum(y2+del)/s2 )
  mu<-rnorm(1,mean.mu,sqrt(var.mu))
  ##update del
  var.del<-  1/(1/t02+ (n1+n2)/s2 )
  mean.del<- var.del*( del0/t02 + sum(y1-mu)/s2 - sum(y2-mu)/s2
  del<-rnorm(1,mean.del,sqrt(var.del))
  ##save parameter values
  MU<-c(MU,mu) ; DEL<-c(DEL,del) ; S2<-c(S2,s2)
  Y12<-rbind(Y12,c(rnorm(2,mu+c(1,-1)*del,sqrt(s2) ) ) )
```

Comparing two groups - Bayesian approach



Comparing two groups - Bayesian approach

```
> quantile(DEL,c(.025,.5,.975))
      2.5%        50%      97.5%
-0.4500081  2.3199015  4.9341240
> quantile(DEL*2,c(.025,.5,.975))
      2.5%        50%      97.5%
-0.9000162  4.6398031  9.8682480
> mean(DEL>0)
[1] 0.95
> mean(Y12[,1]>Y12[,2])
[1] 0.6232
```

Comparing multiple groups

Hierarchical/Multilevel Data
- data where there is a natural grouping structure or hierarchy of nested populations
- Examples?
- we are interested in both within group and between group variability

Comparing multiple groups

Hierarchical Normal Model

Hierarchical data $\{\mathbf{Y}_1 ... \mathbf{Y}_2\}$ where $\mathbf{Y}_j = \left\{Y_{1,j}, ..., Y_{n_j,j}\right\}$ (a random sample from group j, $j = 1, ..., m$).

within group j

$$y_{1,j}, ..., y_{n_j,j} | \theta_j, \sigma^2 \overset{\text{iid}}{\sim} \text{normal}(\theta_j, \sigma^2) \text{ (within group model)}$$

$$\theta_1, ..., \theta_m | \mu, \tau^2 \overset{\text{iid}}{\sim} \text{normal}(\mu, \tau^2) \text{ (between group model)}$$

population mean

between group variance.

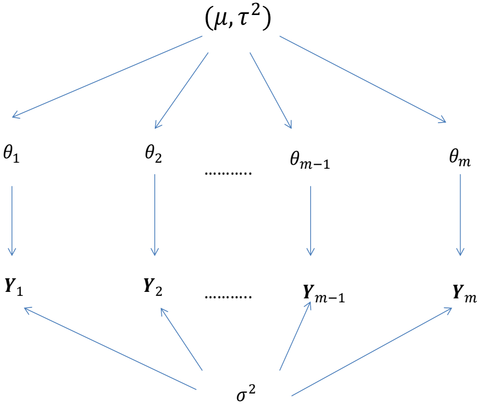
starting value (mean of mean ?)

just use

var($\bar{y}$)

Comparing multiple groups

Hierarchical Normal Model



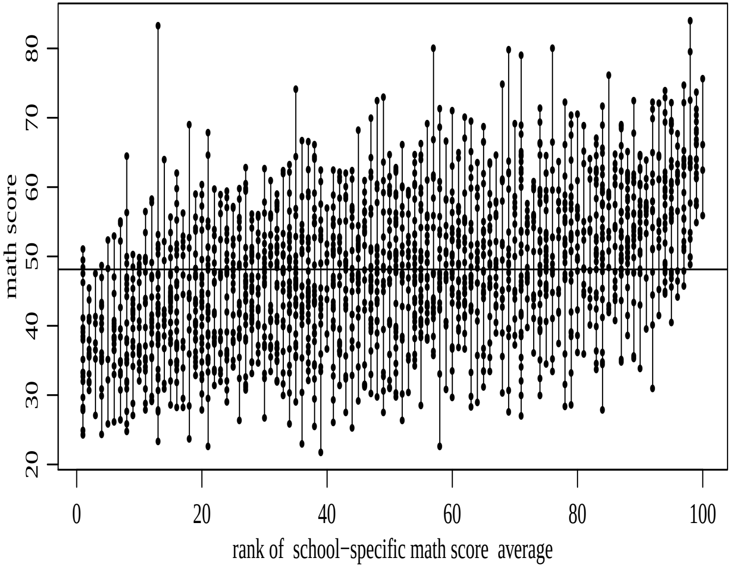
Comparing multiple groups

Hierarchical Normal Model

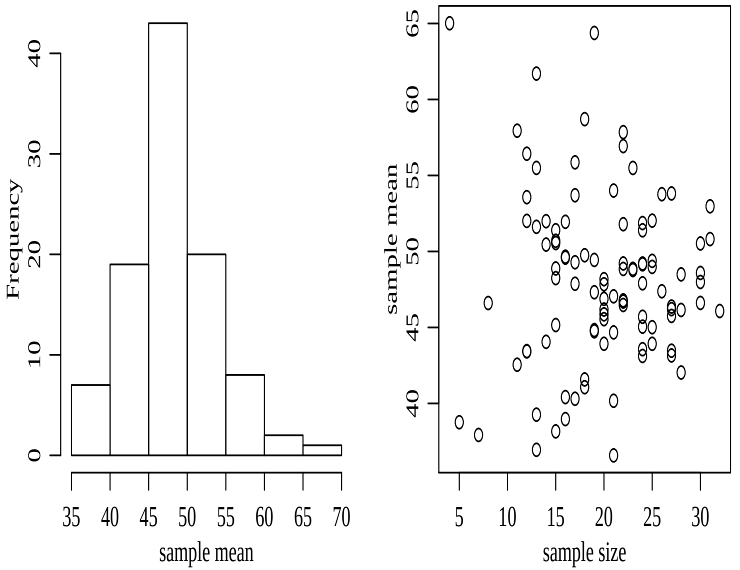
$$1/\sigma^2 \sim \mathrm{gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$
$$1/\tau^2 \sim \mathrm{gamma}(\eta_0/2, \eta_0\tau_0^2/2)$$
$$\mu \sim \mathrm{normal}(\mu_0, \gamma_0^2)$$

Derive the posterior distributions.

Comparing multiple groups - Math scores example



Comparing multiple groups - Math scores example



Comparing multiple groups - Math scores example

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

### starting values

m<-length(unique(school))
n<-sv<-ybar<-rep(NA,m)
for(j in 1:m)
{
  ybar[j]<-mean(mathscore[school==j])
  sv[j]<-var(mathscore[school==j])
  n[j]<-length(mathscore[school==j])
}
theta<-ybar
sigma2<-mean(sv)
mu<-mean(theta)
tau2<-var(theta)
###
```

Comparing multiple groups - Math scores example

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

mu, sigma, theta.

theta has different length in fact.
but for each iteration,
only 1 mu,
1 sigma
1 theta
are generated.

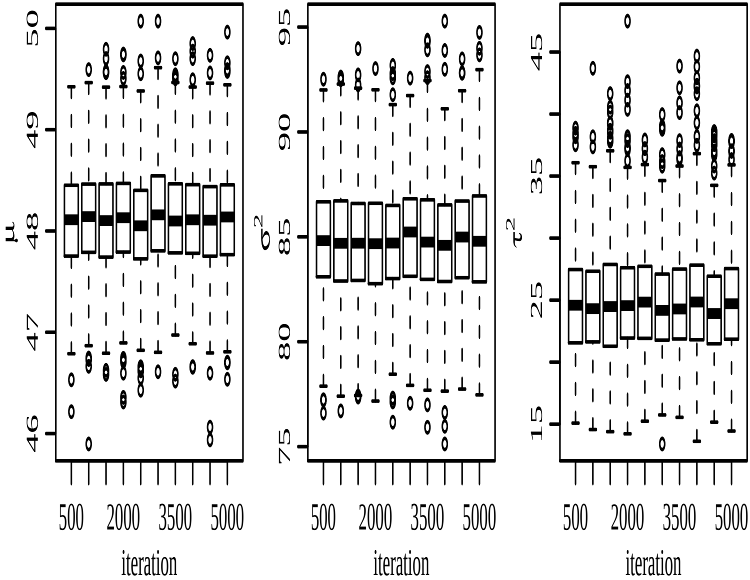
```
### MCMC algorithm
for(s in 1:S)
{

  # 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))
   }
```

Comparing multiple groups - Math scores example

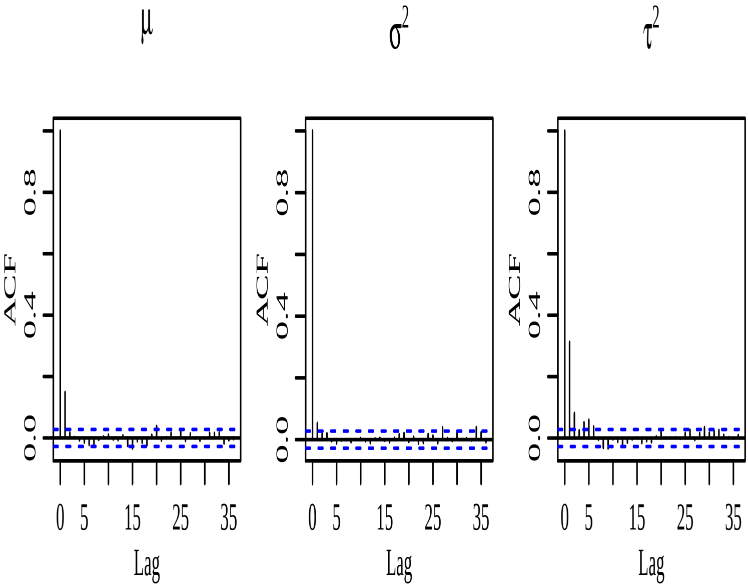
```
#sample new value of sigma2
 nun<-nu0+sum(n)
 ss<-nu0*s20;for(j in 1:m){ss<-ss+sum((mathscore[school==j]
           -theta[j])^2)}
 sigma2<-1/rgamma(1,nun/2,ss/2)

 #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
 MST[s,]<-c(mu,sigma2,tau2)
}
```

Comparing multiple groups - Math scores example

each box means 500 iters diagnostic tracks

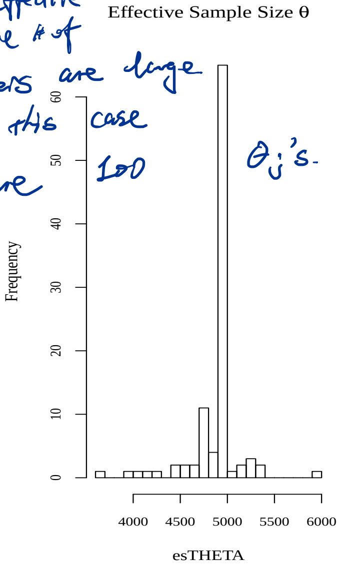


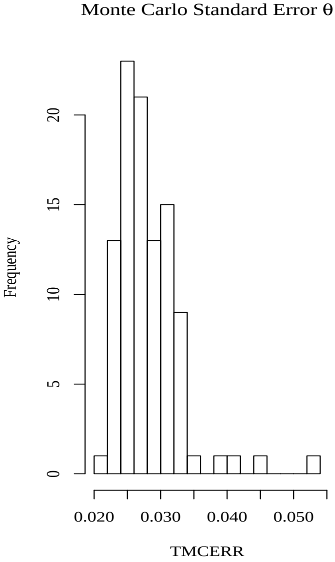
Comparing multiple groups - Math scores example



Comparing multiple groups - Math scores example

effect sample size is more effective when the # of parameters are large e.g. in this case there are 100 θ_j's.



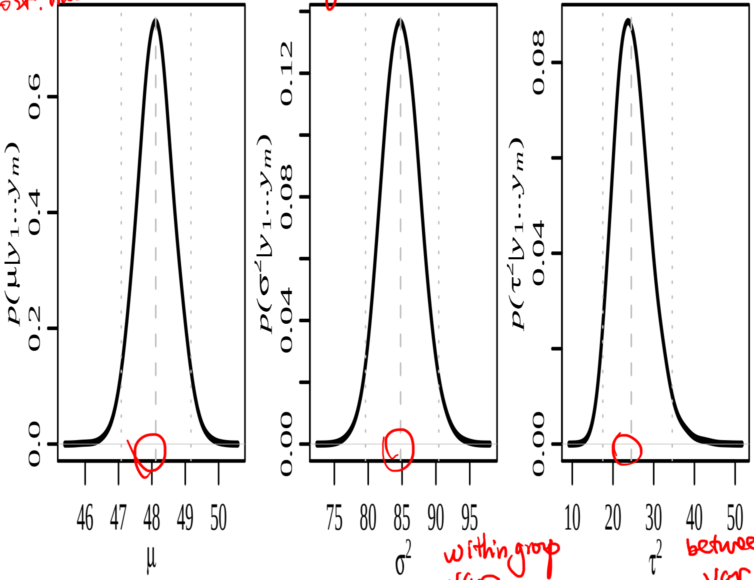


Comparing multiple groups - Math scores example

Posterior Summaries

marginal dist for μ, σ^2, τ^2

post. mode



within group var

between group var

Comparing multiple groups - Math scores example

Shrinkage

$$E[\theta_j | y_j, \mu, \tau, \sigma] = \frac{\bar{y}_j n_j/\sigma^2 + \mu/\tau^2}{n_j/\sigma^2 + 1/\tau^2}$$

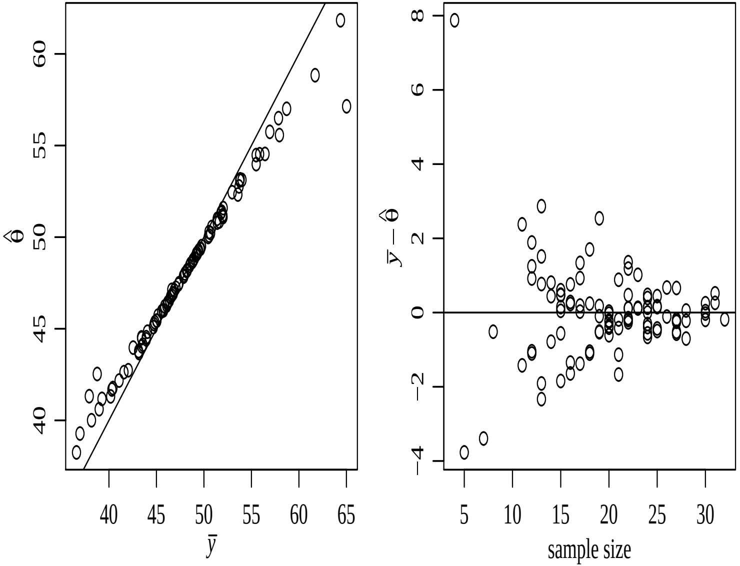
$$= \frac{n_j/\sigma^2}{n_j/\sigma^2 + \frac{1}{\tau^2}} \cdot \bar{y} + \frac{1/\tau^2}{n_j/\sigma^2 + \frac{1}{\tau^2}} \mu$$

no shrinkage $\Rightarrow$ no info shared between groups. groups.

when $n_j \downarrow$, shrinkage $\uparrow$. sample size of j, how to find posterior info of group j? Borrow info from other

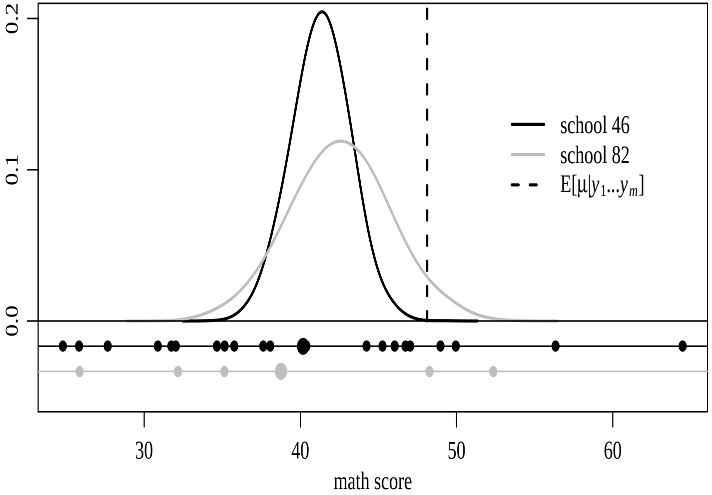
Comparing multiple groups - Math scores example

Shrinkage



Comparing multiple groups - Math scores example

Shrinkage



$$E[\theta_{46}|y_{46},\mu,\tau,\sigma] = 41.31;\ \bar{y}_{46} = 40.18$$
$$E[\theta_{82}|y_{82},\mu,\tau,\sigma] = 42.53;\ \bar{y}_{82} = 38.76$$

Hierarchical modelling of means and variances

Let's allow the sampling variance to vary across groups.

$$Y_{1,j}, ..., y_{n_j,j} \overset{\text{iid}}{\sim} \text{normal}(\theta_j, \sigma_j^2)$$

and so

WHAT'S CHANGED

replace every σ^2 with σ_j^2

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

Also

$$\sigma_1^2, ..., \sigma_m^2 \overset{\text{iid}}{\sim} \text{inv-gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$

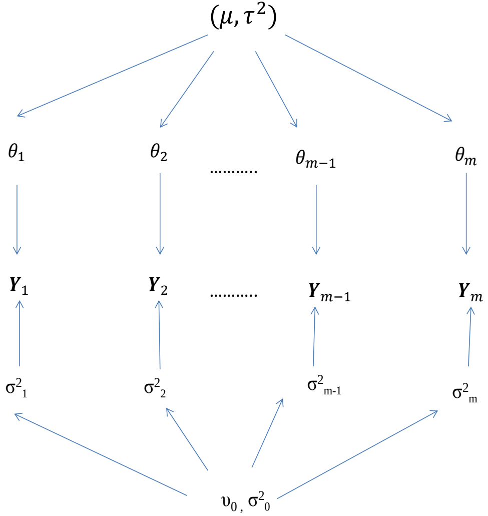
and so

still follows InvGamma.

but we need to consider ν_0 & σ_0^2 for σ_j^2 (new params)

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

Hierarchical modelling of means and variances



another layer of σ^2

Hierarchical modelling of means and variances

Prior density for ν_0 and σ_0^2

Inv Gamma

If $p(\sigma_0^2) \sim \mathrm{gamma}(a, b)$, then we can show that

actually it's gamma, not ~~Inv Gamma~~

$$p(\sigma_0^2|\sigma_1^2, ..., \sigma_m^2, \nu_0) \sim \mathrm{gamma}(a + \frac{1}{2}m\nu_0, b + \frac{\nu_0}{2}\sum_{j=1}^{m}(1/\sigma_j^2))$$

Let the prior density on ν_0 be the geometric distribution so that $p(\nu_0) \propto e^{-\alpha\nu_0}$, then

$$\begin{aligned} p(\nu_0|\sigma_0^2, \sigma_1^2, ..., \sigma_m^2) &\propto p(\nu_0) \times p(\sigma_1^2, ..., \sigma_m^2|\nu_0, \sigma_0^2) \\ &\propto \left(\frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right)^m \times exp\left\{-\nu_0(\alpha + \frac{1}{2}\sigma_0^2\sum_{j=1}^{m}(1/\sigma_j^2))\right\} \end{aligned}$$

How would you sample from this unnormalized distribution?

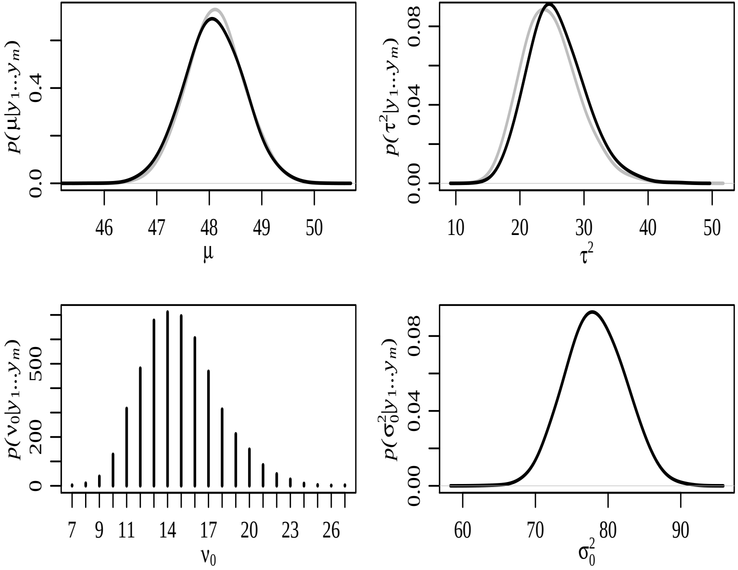
Hierarchical modelling of means and variances

R code to get posterior samples of ν_0.

```
nu0s<-1:5000
#sample new nu0
  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)) )
```

[nb: substracting off the maximum value before exponentiating creates an unnormalised discrete approximation with maximum value 1. The `sample` function in R automatically reweights the values specified in `prob=""` to sum to 1.]

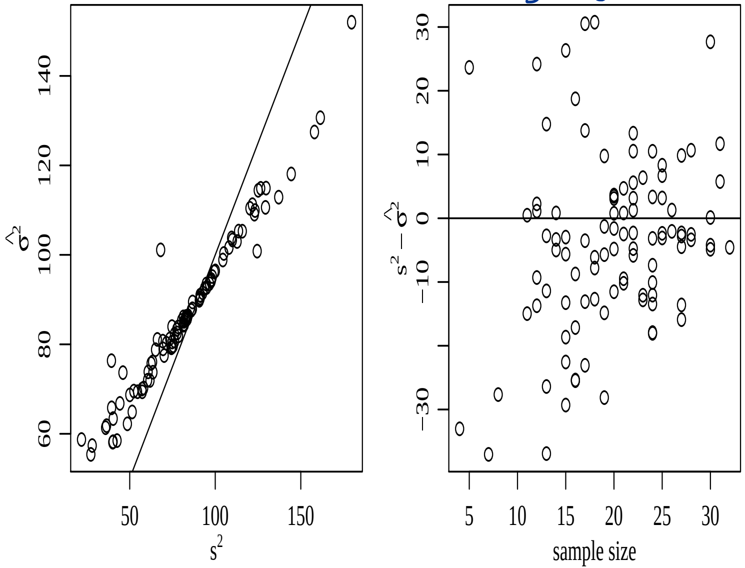
Hierarchical modelling of means and variances



Hierarchical modelling of means and variances

sample large ↑

$s^2 \to \sigma^2$



Exercise - heart transplant mortality data

Data is recorded on the number of deaths within 30 days of heart transplant surgery is recorded for each of 94 hospitals. The data are in the file `hearttransplants.csv` . The variable "y" is the number of deaths and the variable "e" is the exposure. We wish to estimate the mortality rate λ_i per unit of exposure for each hospital $i = 1, ..., 94$.

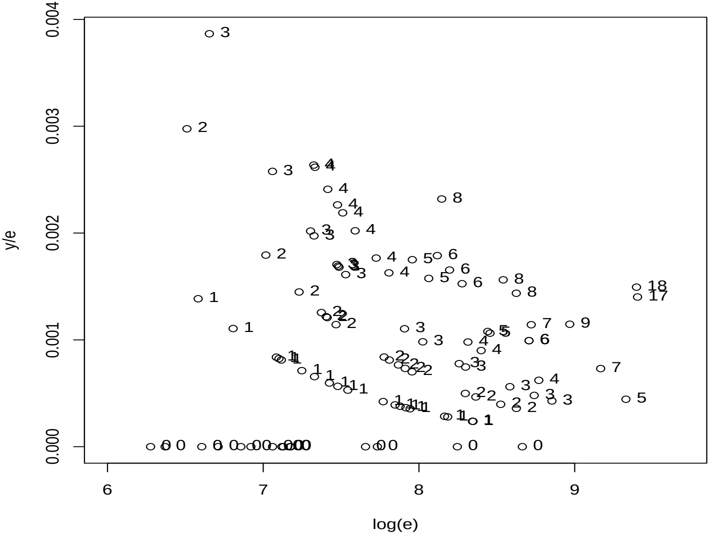
Q: One option is to model the individual mortality rates separately using data for that hospital only. What are some issues with this approach? separate-means model.

equal-means model. (1 λ)

Q: A second option is the equal-means approach, where all the data from all hospitals are pooled together and a single mortality rate estimate is obtained. What are the issues with this approach?

Fit the separate-means and equal-means models and show that the resulting inference is inadequate.

Exericse - heart transplant mortality data



Exercise - heart transplant mortality data

Our aim is to fit Bayesian hierarchical model to estimate the true mortality rates for each hospital. Be sure to answer the following questions:

- What assumptions would you make for the hyperprior, prior and sampling model distributions?
- What is the posterior distribution for mortality rates for each hospital? Discuss how you would obtain posterior draws of the mortality rates from this distribution.
- How would you estimate the shrinkage effect for each hospital?
- Suppose you had to choose a hospital for heart transplant surgery. How would you use your Bayesian analysis to select a hospital?