Model Estimation, Testing, and Reporting

PSYC 575

Mark Lai

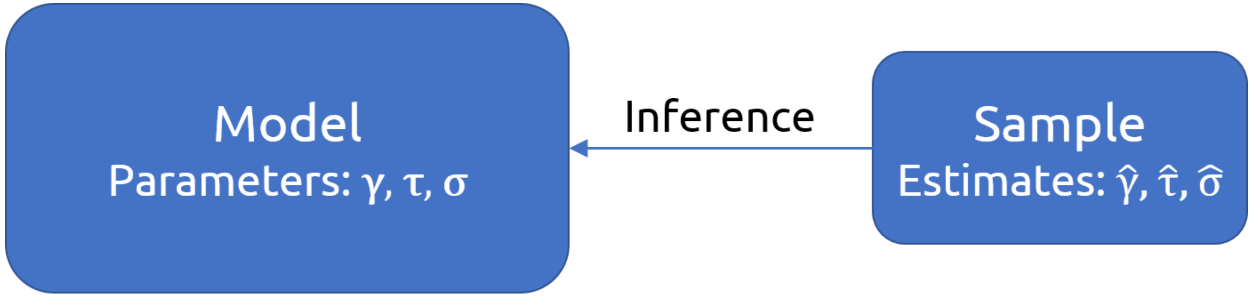
University of Southern California

2020/09/01 (updated: 2020-09-05)

Week Learning Objectives

- Describe conceptually what **likelihood function** and maximum likelihood estimation are
- Describe the differences between **maximum likelihood** and **restricted maximum likelihood**
- Conduct statistical tests for fixed effects
- Use the **likelihood ratio test** to test random slopes
- Report results of a multilevel analysis based on established guidelines

Estimation



Regression: OLS

MLM: Maximum likelihood, Bayesian

Why should I learn about estimation methods?

- Understand software options
- Know when to use better methods
- Needed for reporting

Maximum Likelihood Estimation

The most commonly used methods in MLM are

maximum likelihood (ML) and restricted maximum likelihood (REML)

```
># Linear mixed model fit by REML ['lmerMod']
># Formula: Reaction ~ Days + (Days | Subject)
>#    Data: sleepstudy
># REML criterion at convergence: 1743.628
># Random effects:
>#  Groups   Name        Std.Dev. Corr
>#  Subject  (Intercept) 24.741
>#           Days         5.922   0.07
>#  Residual             25.592
># Number of obs: 180, groups:  Subject, 18
># Fixed Effects:
># (Intercept)         Days
>#      251.41        10.47
```

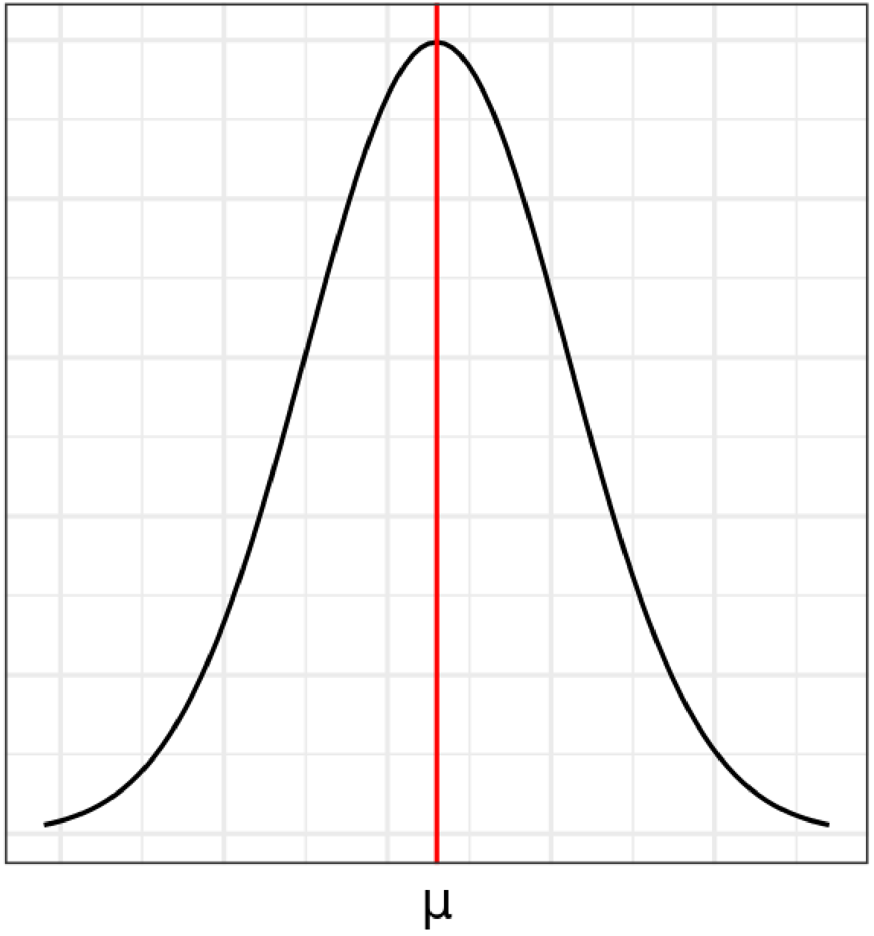
But what is "Likelihood"?

Likelihood

Let's say we want to estimate the population mean math achievement score (μ)

We need to make some assumptions:

- Known *SD*: $\sigma = 8$
- The scores are normally distributed in the population



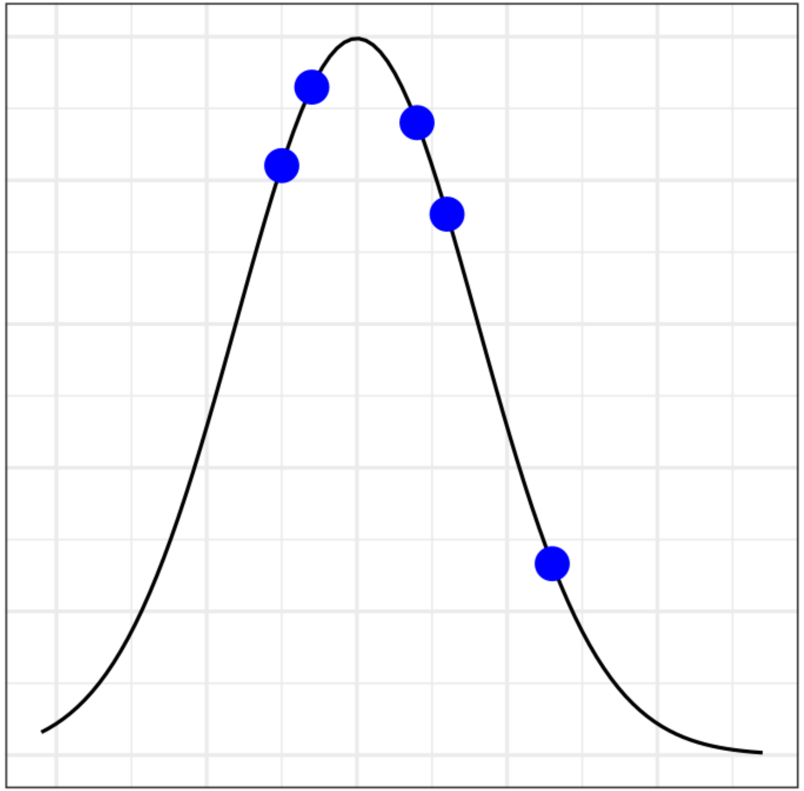
Learning the Parameter From the Sample

Assume that we have scores from 5 representative students

Student	Score
1	23
2	16
3	5
4	14
5	7

Likelihood

If we **assume** that $\mu = 10$, how likely will we get 5 students with these scores?



Student	Score	$P(Y_i = y_i \mid \mu = 10)$
1	23	0.0133173
2	16	0.0376422
3	5	0.0410201
4	14	0.0440082
5	7	0.0464819

Multiplying them all together:

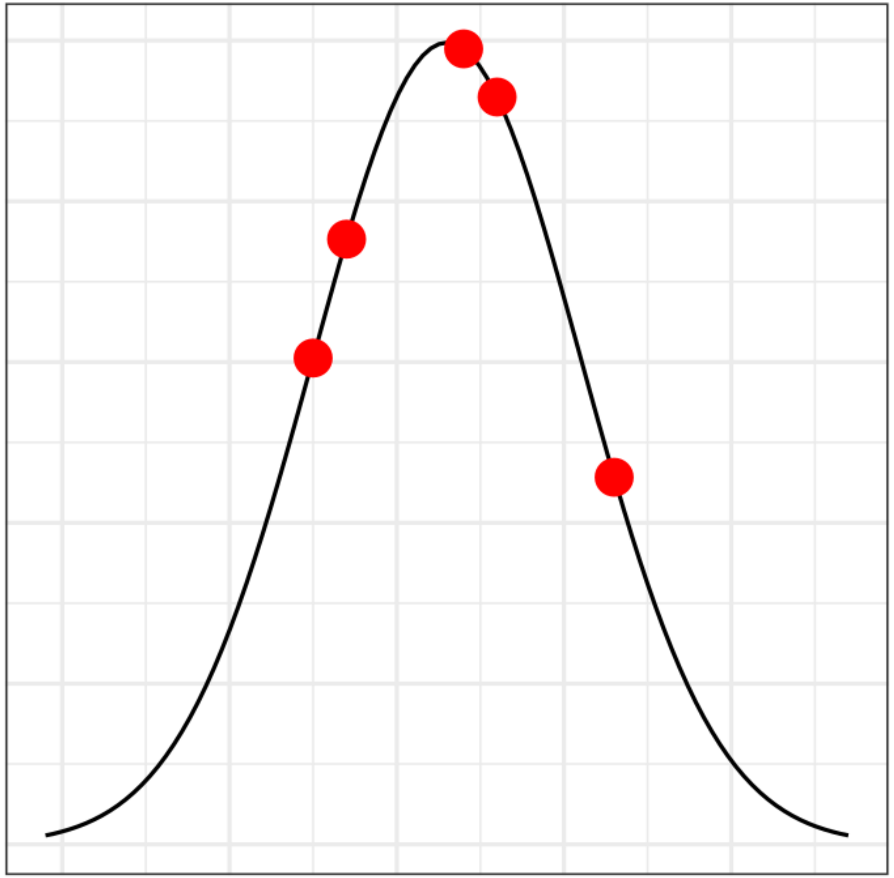
$$P(Y_1 = 23, Y_2 = 16, Y_3 = 5, Y_4 = 14, Y_5 = 7 | \mu = 10)$$

= Product of the probabilities =

```
prod(dnorm(c(23, 16, 5, 14, 7), mean = 10, s
```

```
># [1] 4.20634e-08
```

If $\mu = 13$



Student	Score	$P(Y_i = y_i \mid \mu = 13)$
1	23	0.0228311
2	16	0.0464819
3	5	0.0302463
4	14	0.0494797
5	7	0.0376422

Multiplying them all together:

$P(Y_1 = 23, Y_2 = 16, Y_3 = 5, Y_4 = 14, Y_5 = 7 | \mu = 13)$

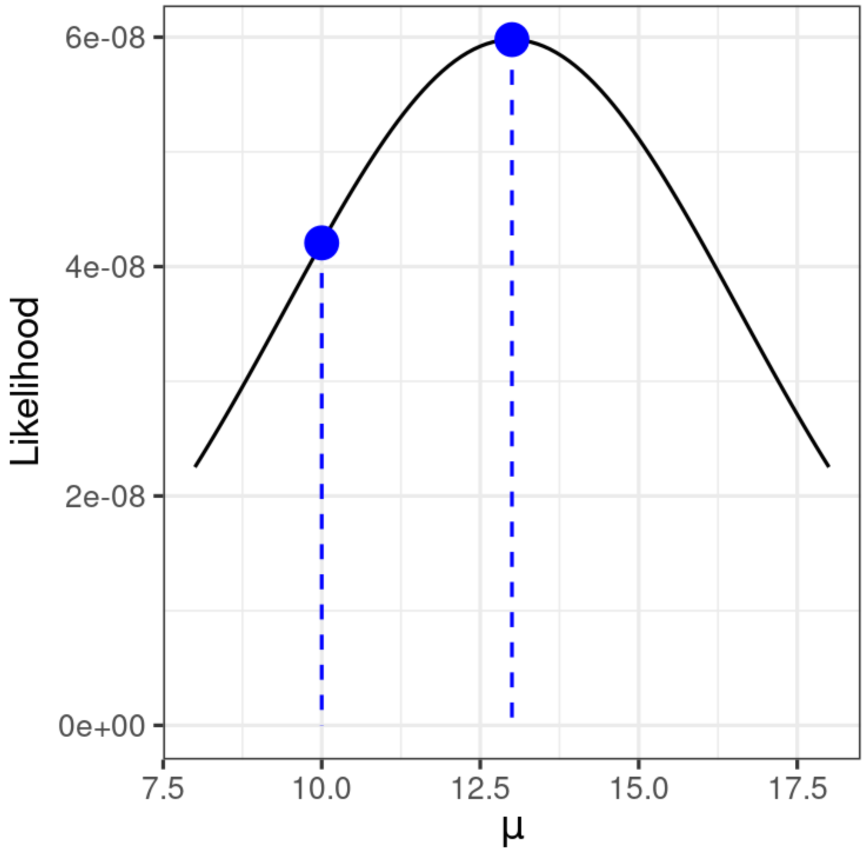
= Product of the probabilities =

```
prod(dnorm(c(23, 16, 5, 14, 7), mean = 13, s
```

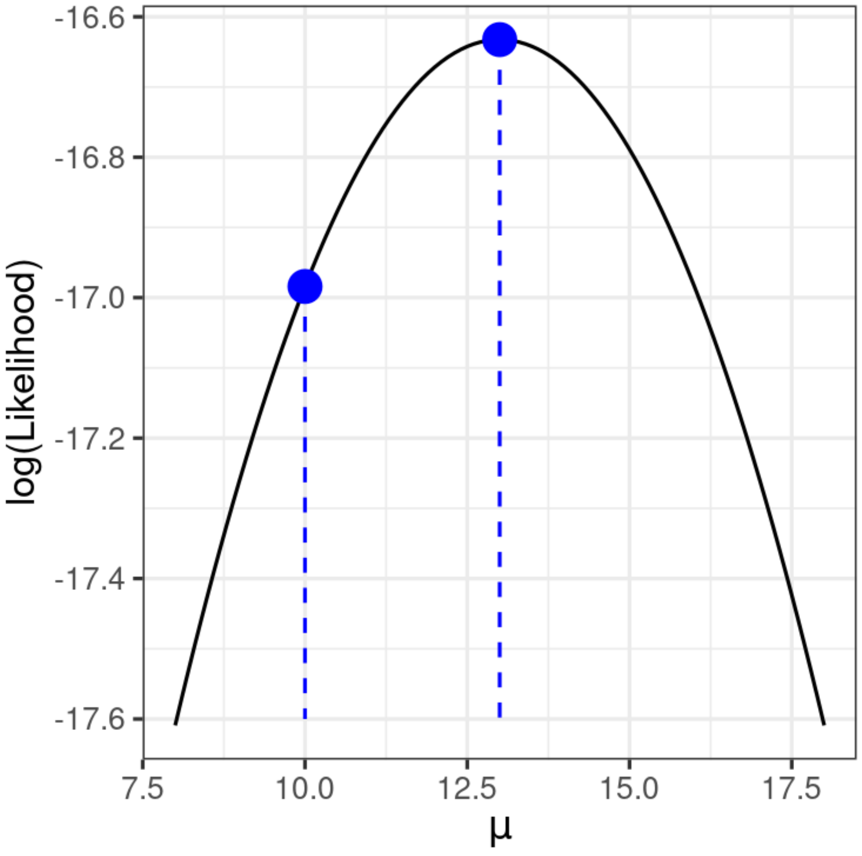
```
># [1] 5.978414e-08
```

Compute the likelihood for a range of μ values

Likelihood Function



Log-Likelihood (LL) Function

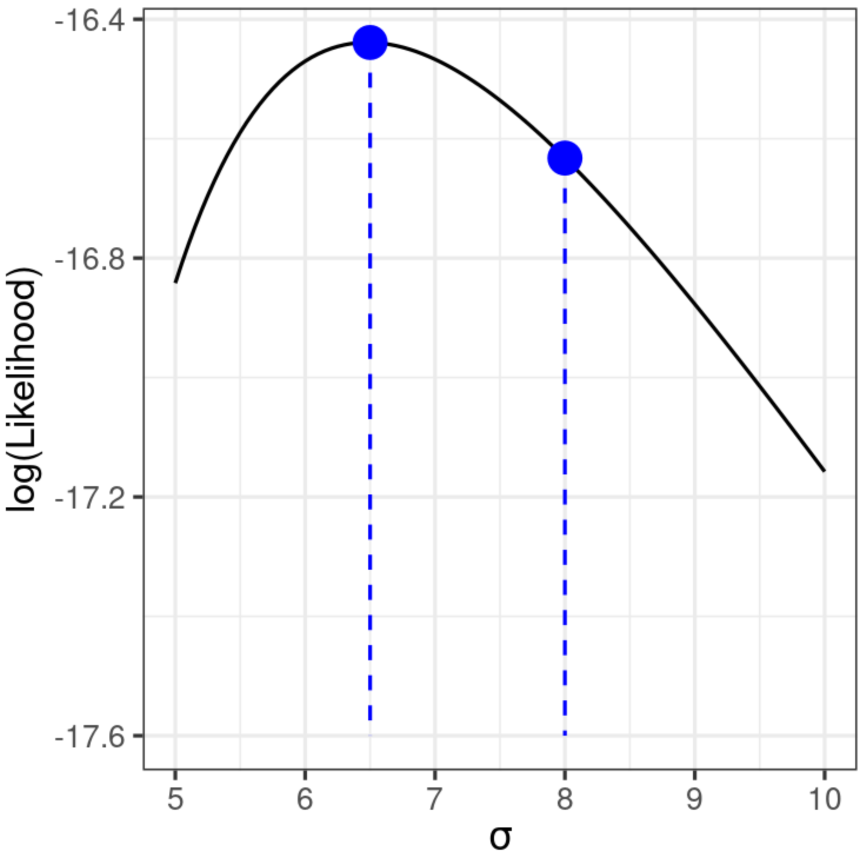


Maximum Likelihood

$\hat{\mu} = 13$ maximizes the (log) likelihood function

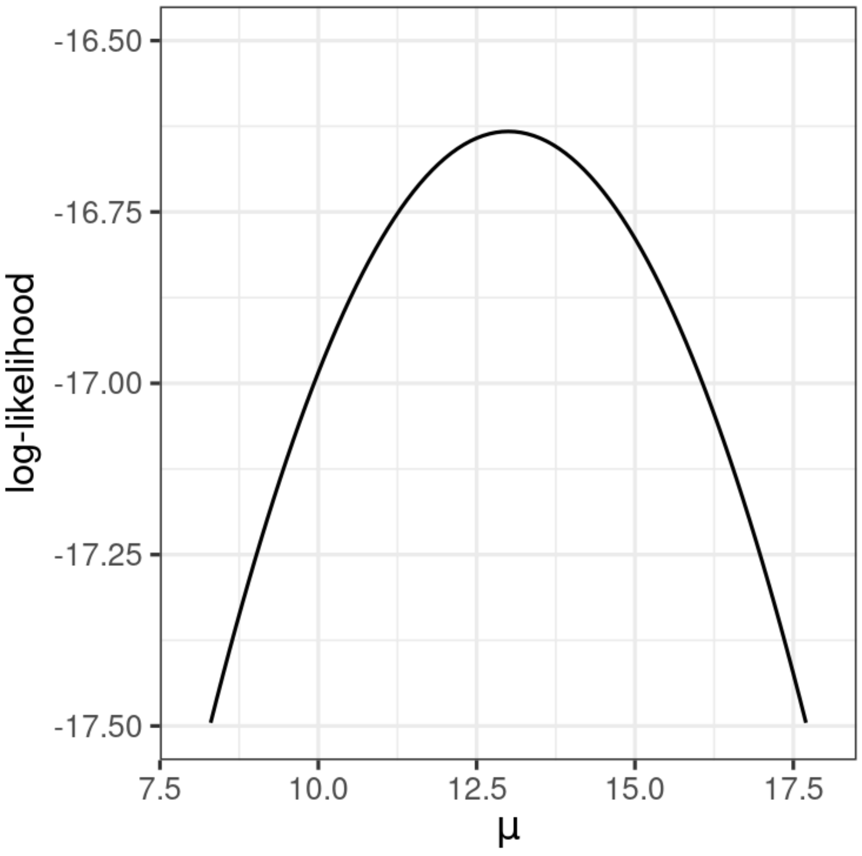
Maximum likelihood estimator (MLE)

Estimating σ

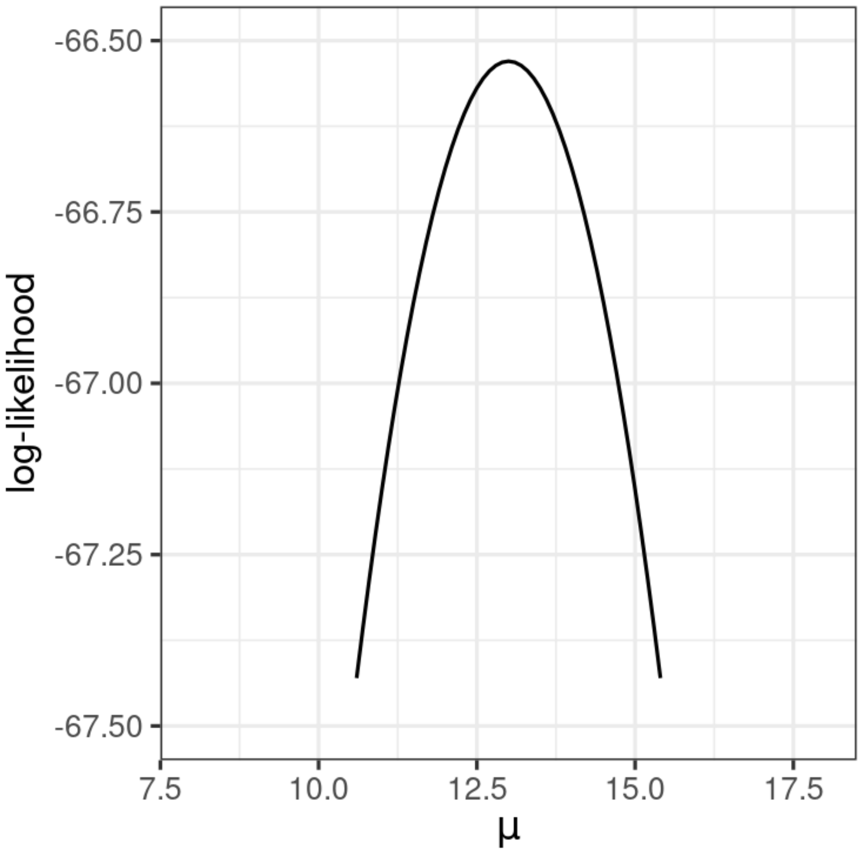


Curvature and Standard Errors

$N = 5$



$N = 20$



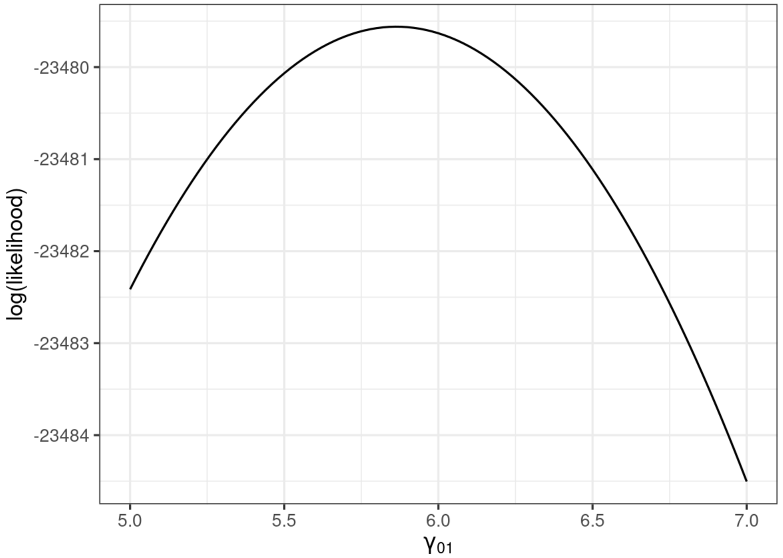
Estimation Methods for MLM

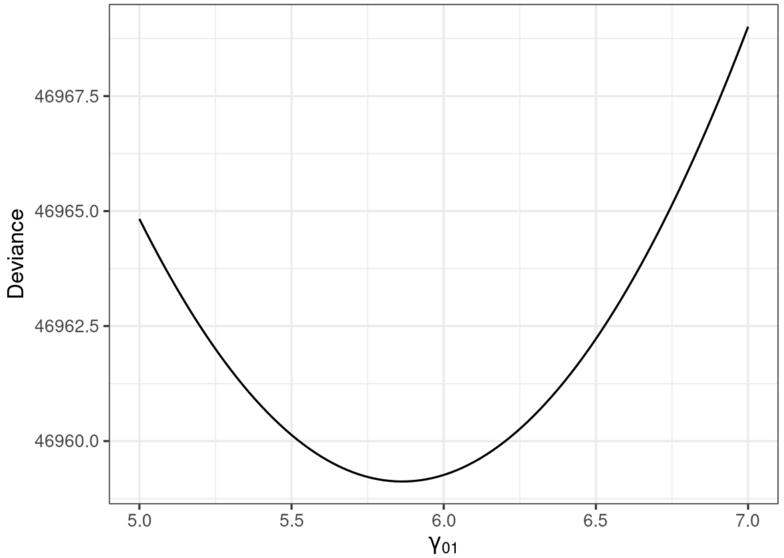
For MLM

Find γs, τs, and σ that maximizes the likelihood function

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\tau}, \sigma; \mathbf{y}) = -\frac{1}{2}\left\{\log|\mathbf{V}(\boldsymbol{\tau}, \sigma)| + (\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})^\top \mathbf{V}^{-1}(\boldsymbol{\tau}, \sigma)(\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})\right\} + K$$

Here's the log-likelihood function for the coefficient of `meanses` (see code in the provided Rmd):

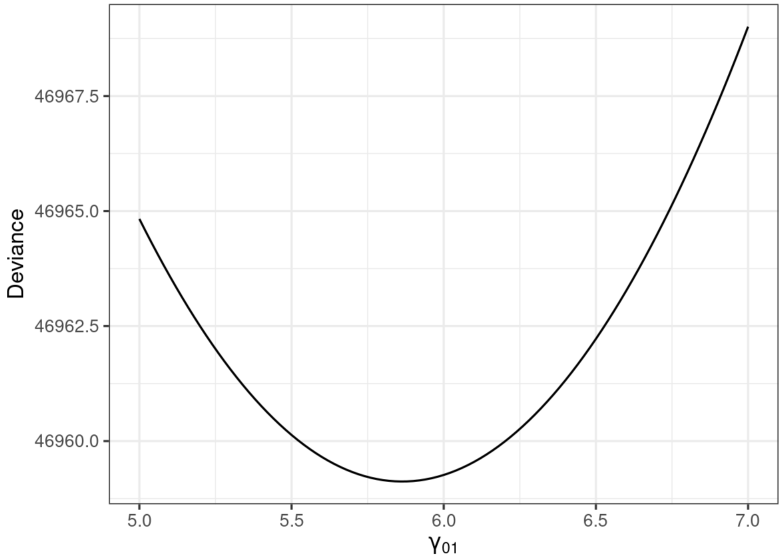




Numerical Algorithms

```r
m_lv2 <- lmer(mathach ~ meanses + (1 | id),
```

```
>#  iteration: 1
>#      f(x) = 47022.519159
>#  iteration: 2
>#      f(x) = 47151.291766
>#  iteration: 3
>#      f(x) = 47039.480137
>#  iteration: 4
>#      f(x) = 46974.909593
>#  iteration: 5
>#      f(x) = 46990.872588
>#  iteration: 6
>#      f(x) = 46966.453125
>#  iteration: 7
>#      f(x) = 46961.719993
>#  iteration: 8
>#      f(x) = 46965.890703
>#  iteration: 9
>#      f(x) = 46961.367013
>#  iteration: 10
```



ML vs. REML

REML has corrected degrees of freedom for the variance component estimates (like dividing by $N - 1$ instead of by N in estimating variance)

- REML is generally preferred in smaller samples
- The difference is small with large number of clusters

Technically, REML only estimates the variance components[1]

[1] The fixed effects are integrated out and are not part of the likelihood function. They are solved in a second step, usually by the generalized least squares (GLS) method

160 Schools

	REML	ML
(Intercept)	12.649	12.650
	(0.149)	(0.148)
meanses	5.864	5.863
	(0.361)	(0.359)
sd__(Intercept)	1.624	1.610
sd__Observation	6.258	6.258
AIC	46969.3	46967.1
BIC	46996.8	46994.6
Log.Lik.	-23480.642	-23479.554
REMLcrit	46961.285	

16 Schools

	REML	ML
(Intercept)	12.809	12.808
	(0.504)	(0.471)
meanses	6.577	6.568
	(1.281)	(1.197)
sd__(Intercept)	1.726	1.581
sd__Observation	5.944	5.944
AIC	4419.6	4422.2
BIC	4437.7	4440.3
Log.Lik.	-2205.796	-2207.099
REMLcrit	4411.591	

Other Estimation Methods

Generalized estimating equations (GEE)

- Robust to some misspecification and non-normality
- Maybe inefficient in small samples (i.e., with lower power)
- See Snijders & Bosker 12.2; the `geepack` R package

Markov Chain Monte Carlo (MCMC)/Bayesian

Testing

Fixed effects (γ)

- Usually the likelihood-based CI/likelihood-ratio (LRT; χ^2) test is sufficient
- Small sample (10--50 clusters): Kenward-Roger approximation of degrees of freedom
- Non-normality: Residual bootstrap[1]

Random effects (τ)

- LRT (with p values divided by 2)

[1]: See van der Leeden et al. (2008) and Lai (2020)

Likelihood Ratio (Deviance) Test

$$H_0 : \gamma = 0$$

Likelihood ratio: $\dfrac{L(\gamma = 0)}{L(\gamma = \hat{\gamma})}$

Deviance: $-2 \times \log\left(\frac{L(\gamma=0)}{L(\gamma=\hat{\gamma})}\right)$
= $-2\text{LL}(\gamma = 0) - [-2\text{LL}(\gamma = \hat{\gamma})]$
= $\text{Deviance}\ |_{\gamma=0}\ - \text{Deviance}|_{\gamma=\hat{\gamma}}$

ML (instead of REML) should be used

Example

```
...
>#  Linear mixed model fit by maximum likelihood  ['lmerMod']
>#  Formula: mathach ~ (1 | id)
>#        AIC       BIC    logLik  deviance  df.resid
>#   47121.81  47142.45 -23557.91  47115.81      7182
...
```

```
...
>#  Linear mixed model fit by maximum likelihood  ['lmerMod']
>#  Formula: mathach ~ meanses + (1 | id)
>#        AIC       BIC    logLik  deviance  df.resid
>#   46967.11  46994.63 -23479.55  46959.11      7181
...
```

```r
pchisq(47115.81 - 46959.11, df = 1, lower.tail = FALSE)
```

```
>#  [1] 5.952567e-36
```

In `lme4`, use

```r
drop1(m_lv2, test = "Chisq")  # Automatically use ML
```

F Test With Small-Sample Correction

Needs approximation of degrees of freedom (*df*)

Kenward-Roger approximation generally performs well with < 50 clusters

```
library(lmerTest)
m_contextual <- lmer(mathach ~ meanses + ses + (1 | id),
                     data = hsbsub)
anova(m_contextual, ddf = "Kenward-Roger")
```

```
># Type III Analysis of Variance Table with Kenward-Roger's method
>#          Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)    
># meanses  324.39  324.39     1  15.51  9.9573  0.006317 ** 
># ses     1874.34 1874.34     1 669.03 57.5331 1.116e-13 ***
># ---
># Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

LRT for Random Slopes

Should you include random slopes?

Theoretically yes unless you're certain that the slopes are the same for every groups

However, frequentist methods usually crash with more than two random slopes

- Test the random slopes one by one, and identify which one is needed
- Bayesian methods are more equipped for complex models

"One-tailed" LRT

LRT (χ^2) is generally a two-tailed test. But for random slopes,

$H_0 : \tau_1 = 0$ is a one-tailed hypothesis

A quick solution is to divide the resulting p by 2[1]

[1]: Originally proposed by Snijders & Bosker; tested in simulation by LaHuis & Ferguson (2009, https://doi.org/10.1177/1094428107308984)

Example: LRT for τ_1^2

```
...
># Formula: mathach ~ meanses + ses_cmc + (ses_cmc | id)
>#    Data: hsball
># REML criterion at convergence: 46557.65
...
```

```
...
># Formula: mathach ~ meanses + ses_cmc + (1 | id)
>#    Data: hsball
># REML criterion at convergence: 46568.58
...
```

G Matrix

$$\begin{bmatrix} \tau_0^2 & \\ \tau_{01} & \tau_1^2 \end{bmatrix}$$

$$\begin{bmatrix} \tau_0^2 & \\ 0 & 0 \end{bmatrix}$$

```r
pchisq(10.92681, df = 2, lower.tail = FALSE)
```

```
># [1] 0.004239097
```

Need to divide by 2

Reporting

McCoach (2019 chapter), see HW 5