S-043/Stat-151
Analysis for Clustered and Longitudinal Data
(Multilevel & Longitudinal Models)

Unit 2, Lecture 6: Inference for MLMs

Instructor: Prof. Luke W. Miratrix
lmiratrix@g.harvard.edu
Larsen 603

xkcd.com


Office hours: a bit under-attended
http://desperateandlonelyanimals.tumblr.com

Final Projects Preliminary Proposal

Due Monday, October 7th (revised date)

Easy and straightforward, low stakes.

We are just starting the conversation.

Our MLM give us three types of things

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right]$$

Estimated fixed effects

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

Estimated variance and covariance parameters

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

With Empirical Bayes *as a second step,* Estimated random effects for each group

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

How do we assess uncertainty?

We have our three things:

- ★ Estimated fixed effects
- ★ Estimates of the variances and covariances of the hyperparameters
- ★ Estimates of the random effects for each group (from Empirical Bayes as a second step)

We have seen how to get **point estimates.**

We also need **uncertainty** for those estimates.

And we need to know how to conduct **inference** for these estimates ("are they significant?")

Today's Goals

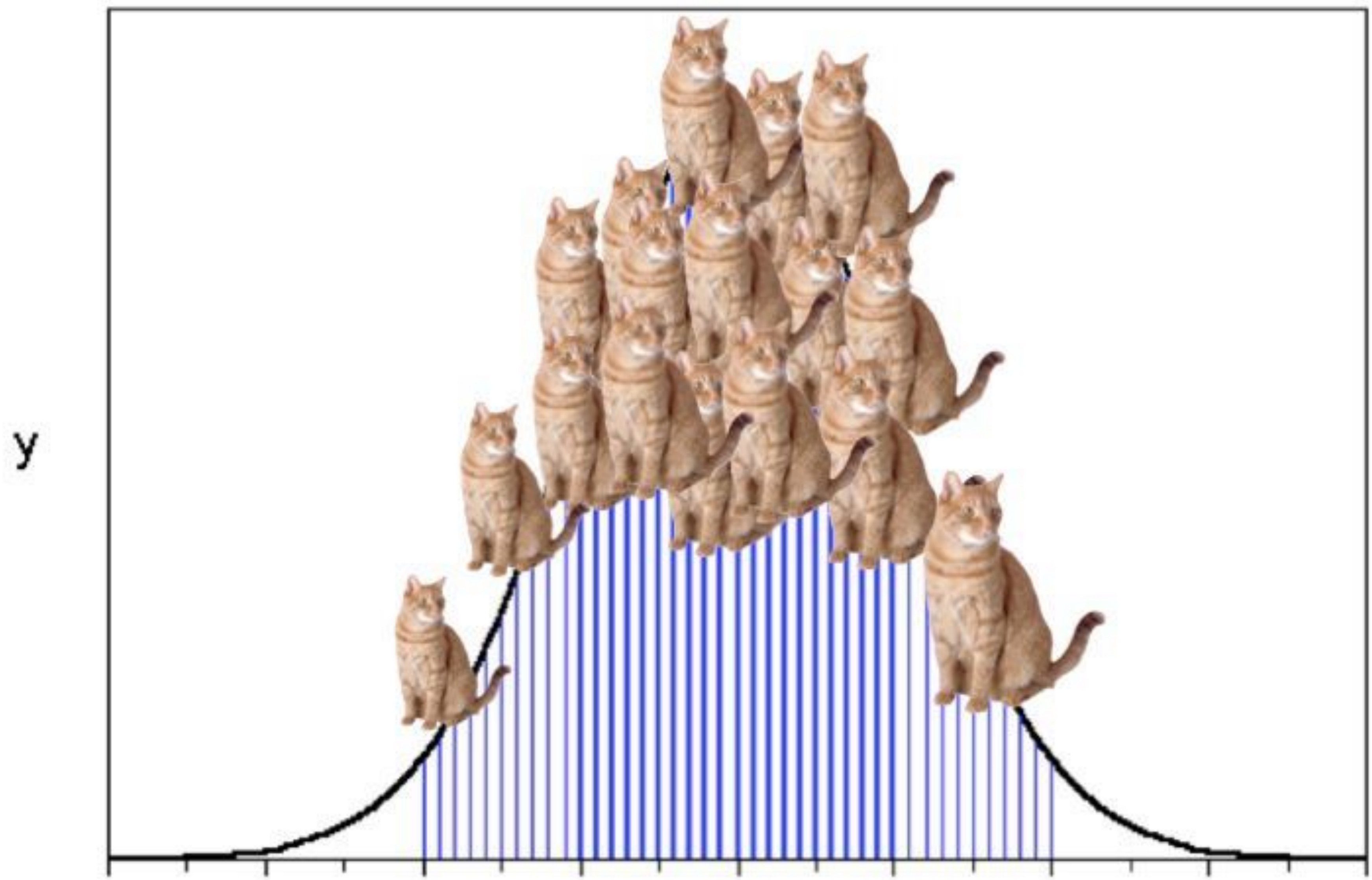
★ Talk about generating confidence intervals
  - We can use the normal approximation or profile intervals

★ Talk about how to do hypothesis testing
  - Discuss and illustrate several different testing procedures

★ Give more intuition for the ideas behind likelihood estimation and testing

★ Continue to identify the different kinds of things we estimate with multilevel models

Standard errors and normal assumption-based confidence intervals



The Normal Approximation

Maximum Likelihood Estimation provides some nice things:

- ★ **Consistency**
  (as the sample grows, uncertainty shrinks towards 0)
- ★ **Asymptotically Normally** distributed point estimates
  (in big samples, point estimates follow a normal distribution across repeated random samples)
- ★ Approximate (asymptotically correct) **standard errors**
  (as the sample grows, our estimates of the standard errors will approach the correct ones)

I.e., roughly speaking, for any estimand β we have:

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to N(0, \tau^2)$$

Or "as the sample grows, the estimated parameters will be normally distributed about the true parameters, and the variance will shrink with the sample size towards an estimable value."

The 95% Confidence Interval

If the sampling distribution is relatively symmetric and bell-shaped, a 95% confidence interval can be estimated using

$$statistic \pm 2 \times \widehat{SE}$$

Sample Language:

"We are 95% confident that the true proportion of all Americans that considered the economy a 'top priority' in January 2012 is between 0.84 and 0.88"

Standard Errors and Confidence Intervals for the Fixed Effects

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

"Fixed Effects": se.fixef()

From the "arm" package

```
> M1 = lmer( mathach ~ ses + (ses|id), data=dat )

> fixef( M1 )
(Intercept)         ses
      12.67        2.39

> se.fixef( M1 )
(Intercept)         ses
      0.190       0.118

> # lower and upper bounds of a 95% confidence interval
(Normal Approximation)
> fixef( M1 ) - 2 * se.fixef( M1 )
(Intercept)         ses
      12.29        2.16
> fixef( M1 ) + 2 * se.fixef( M1 )
(Intercept)         ses
      13.04        2.63
```

Or use critical values from a t-distribution with df calculated by the lmerTest package; may be slightly better, especially if there are few level-2 units.

Confidence Intervals for variance and covariance estimates

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

An unfortunate hiccup: Standard Errors for Hyperparameters

We can get our estimates:

```
> VarCorr( M1 )
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 2.197
          ses         0.643    -0.11
 Residual             6.069
```

But getting uncertainty on these estimates is **hard**, and what you get is **unreliable.**

Generally avoid doing it.

(We will test against them being 0 later on.)

Why are there no standard errors for the variance parameters?

From the creator of the lme4 package:

a) “because they are awkward to calculate”

and

b) “because I don’t think they make sense”

He then goes on to say:

“Quoting an estimate and a standard error of the estimate for a parameter is useful if you think that the distribution of the estimator is more-or-less symmetric.”

Variance parameters tend to have skewed distributions, with lots of samples having estimates which are slightly below the true value, and a handful having estimates which are far above.

Related issue: bad boundaries

For the variance parameters you will often see statements such as

“parameter estimates are often on the boundary of the parameter space.”

What does this mean?

Why does this matter for our variance estimation issues?

Boundaries distort normality

If a variance *parameter* is equal (or close) to 0, the distribution of the variance *estimate* can't be normal; a normal distribution is symmetric and an estimated variance parameter can't be negative.

This might not seem like a big deal, but it makes standard errors much less meaningful.

When this happens

$$\hat{\sigma}^2 \pm 2\widehat{SE}(\hat{\sigma}^2)$$

Bad when we are close to zero

is not a valid approach here; we'll get meaningless confidence intervals, and they won't have good coverage rates.

But we can get confidence intervals

```
> display( M2 )
lmer(formula = mathach ~ 1 + ses + sector + (1 | id),
data = dat)
            coef.est coef.se
(Intercept) 11.72    0.23
ses          2.37    0.11
sector       2.10    0.34
```

```
> confint( M2 )

              2.5 %  97.5 %
.sig01        1.653   2.194
.sigma        5.986   6.188
(Intercept)  11.272  12.165
ses           2.167   2.588
sector        1.433   2.770
```

! These are *profile likelihood based confidence intervals.*

This is fancy statistics, and basically asks what is likely if we let other parameters fit as best it can. See Appendix.

Inference

or

Come here, you giant mass of statistical theory, let me scritch you.

Types of Tests

★ Likelihood Ratio Tests

- Our bread & butter test. Versatile, flexible, awesome.

★ Wald tests (i.e., t-tests or the tests you have always known and loved)

- Good, but not good for variance parameters.
- Wald is when we square the statistic (chi-squared). t is the usual single parameter test.

★ ANOVA / F Tests

- Boils down to including clusters as a factor (fixed effects)

★ Score test (Lagrange multiplier)

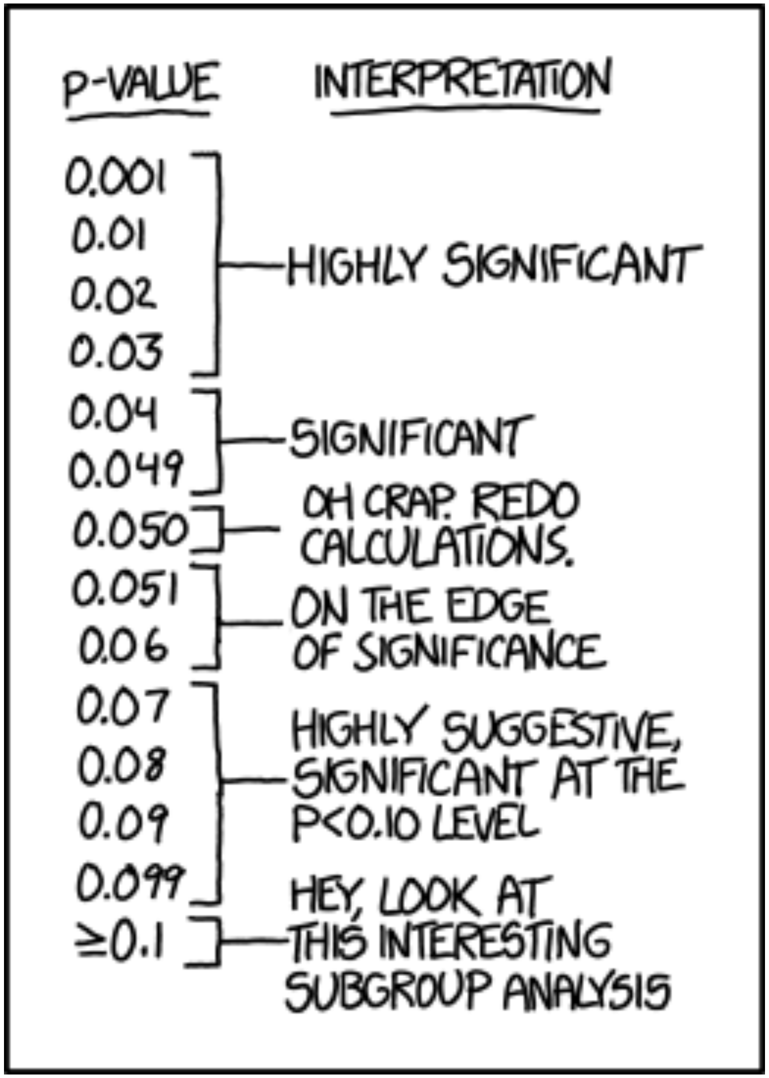
- Not that useful. Ignore.

There is **a lot** of controversy about using many of these tests.

They are mostly based on asymptotic approximations.

E.g., the *t* statistics are not *actually from* a *t* distribution.

P-values and testing



Andrew Gelman "There are no zeros in social science"

My response:

One-sided tests tell you if something is on the "good" side of a null boundary.

Also a reframe: "Rejection means the effect was big enough to notice."

https://xkcd.com/1478/

Inference for fixed effects (this is what you're used to)

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right]$$

Meaning of some hypothesis tests

$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

H_0: $\gamma_{00} = 0$: in the average public school, students with SES of 0 will on average have a math achievement score of 0.

H_0: $\gamma_{01} = 0$: average math achievement for kids with ses=0 in catholic schools is the same as for similar kids in public schools.

H_0: $\gamma_{10} + \gamma_{11} = 0$: in the average Catholic school, there is no association between SES and achievement.

The t-Test

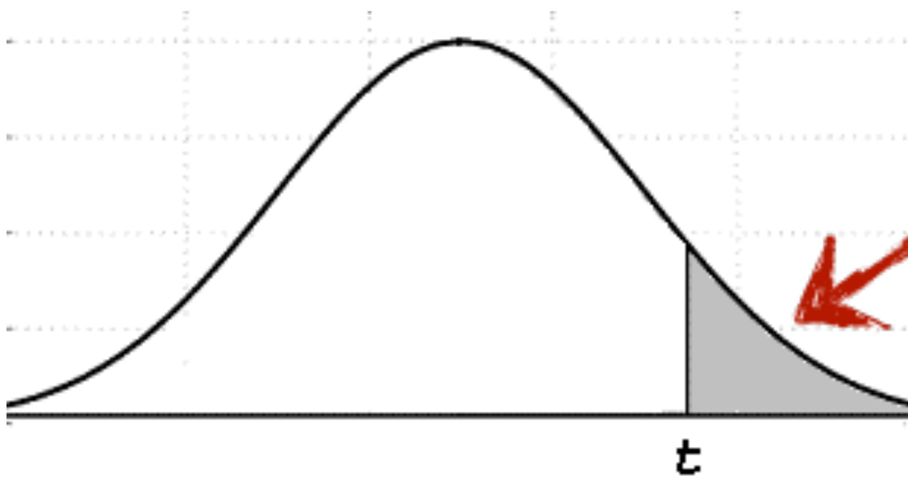
The t statistic:

Get these from the fit model

$$t = \frac{\hat{\beta} - \beta_{hyp}}{\hat{se}(\hat{\beta})}$$

This is usually 0

The distribution

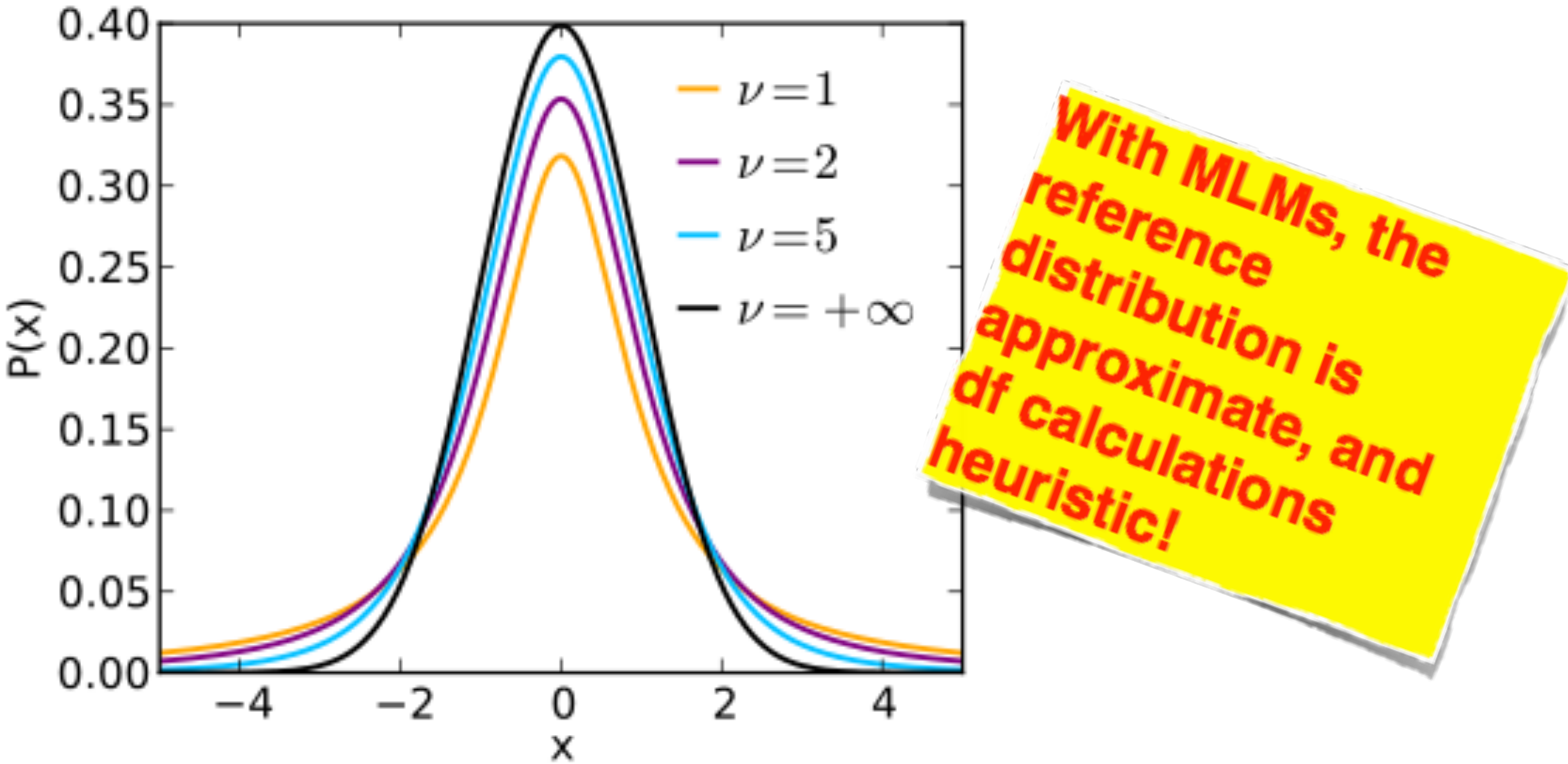


The p-value is the chance of reaching your "extreme value" by chance if the null-hypothesis is correct

(Multiply p by 2 for two-sided tests; note that this assumes a symmetric distribution, which the t-distribution is)

A problem with *t*-tests

The *t*-distribution is indexed by a *degree of freedom (df);* we need this to determine the p-value



Your correct *reference distribution* depends on this number

Image from Wikipedia

Does SES relate to math achievement?

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \epsilon_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Our null hypothesis:

$$H_0 : \gamma_{10} = 0$$

? Can you come up with a better model for testing this hypothesis?

Testing $H_0 : \gamma_{10} = 0$ from our printout

```
> M1 = lmer(formula = mathach ~ 1 + ses + sector +
                (1 + ses | id), data = dat)
> display( M1 )
              coef.est coef.se
(Intercept) 11.47     0.23
ses          2.39     0.12
sector       2.54     0.34

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 1.99
          ses         0.66     0.55
 Residual             6.07
---
number of obs: 7185, groups: id, 160
AIC = 46615.9, DIC = 46593
deviance = 46597.4
```

Estimate of slope between SES and math achievement: $\hat{\gamma}_{01}$

The Estimated Standard Error (SE) for this slope: $\widehat{SE}\,(\hat{\gamma}_{10})$

$$t = \frac{2.39 - 0}{0.12} = 19.9$$

Getting the p-value

```
> t = 2.39 / 0.12
> t
[1] 20.24
    # yikes, that is big!

> 2 * pnorm( t, lower.tail=FALSE )
[1] 2.28e-91          # p-value is basically 0.
```

And as a t-test...

R&B (pg 58) suggest using a df of J - S_q - 1 for level-2 predictors.
J = # Groups
Sq = # covariates at level 2

```
> Sq = 1

> J
[1] 160

> 2 * pt( t, df = J - Sq - 1, lower.tail=FALSE )
[1] 6.146089e-45
```

Automatic p-values

```
> library( lmerTest )
> M1 = lmer( mathach ~ 1 + ses + sector + (1 + ses|id), data=da
> summary( M1 )
Linear mixed model fit by REML t-tests use Satterthwaite
approximations to degrees of freedom

REML criterion at convergence: 46601.9

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept)  3.9645  1.991
          ses          0.4343  0.659    0.55
 Residual             36.8008  6.066
Number of obs: 7185, groups:  id, 160

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  11.4730     0.2315 153.7900  49.568  < 2e-16 ***
ses           2.3854     0.1179 157.8500  20.238  < 2e-16 ***
sector        2.5407     0.3445 151.3100   7.375 1.01e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
```

$$t = \frac{2.385 - 0}{0.118} = 12.2$$

Why so many different df?

Variables which are ostensibly level-1 may really be a mixture of level-1 and level-2

```
Fixed effects:
             Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)    12.094      0.240  196.646   50.36  < 2e-16 ***
female         -1.197      0.164 6581.444   -7.28  3.6e-13 ***
ses             2.349      0.116  159.733   20.25  < 2e-16 ***
sector          2.542      0.334  151.045    7.62  2.6e-12 ***
```

For example:

If SES is mostly explained by school id (students in the same school have very similar values of SES) then while it is a level-1 variable, it has fewer df than something like gender, which is more balanced across school. It’s more like a pseudo-level-2 variable.

The Likelihood Ratio Test

The likelihood ratio test

Fit one model where our target parameter (e.g., γ_{10}) is constrained to be equal to the null-hypothesized value (usually H_0: $\gamma_{10} = 0$) and see how likely the data are. Call this likelihood L_0

Fit another model where γ_{10} is estimated freely, and see how likely the data are *now*. Call this L_1

Note that we need the actual likelihoods, so we need to fit the models with REML = FALSE (although setting REML = TRUE will generally give very similar results)

Restricted Maximum Likelihood (REML) vs (Full) Maximum Likelihood (ML)

For ML, we estimate variances and covariances conditional on the point estimates of the fixed effects.

We end up with overly optimistic (too small) variance estimates.

For REML vs ML, we see most differences in the estimates of the second-level covariance matrix

Shrinkage in variance estimates of roughly (J-F) / J with F being number of level-2 predictors, J number of groups

R&B pages 52-53

Back to Likelihood Testing: A Concept check

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$H_0 : \gamma_{10} = 0$$

Suppose that H_0 is false (γ_{10} is not 0). What do you think will be true of L_0 and L_1?

Now suppose that H_0 is true (γ_{10} *is* 0). What do you think will be true of L_0 and L_1?

Logic of the likelihood ratio test

Take away:

If the likelihoods are basically the same, we accept the null. If the *constrained* model fits like C**P, then we reject the null.

To test we compare the *ratio of the likelihoods*.

Basic idea: if γ_{00} is actually 0, allowing it to be estimated shouldn't change the model fit much.

(I lied in the above: Instead of looking at a ratio of likelihoods, we actually compare a difference of log-likelihoods. But this is the same thing. Why?)

Some Intuition for Likelihood Ratio Tests

Likelihood tests ask

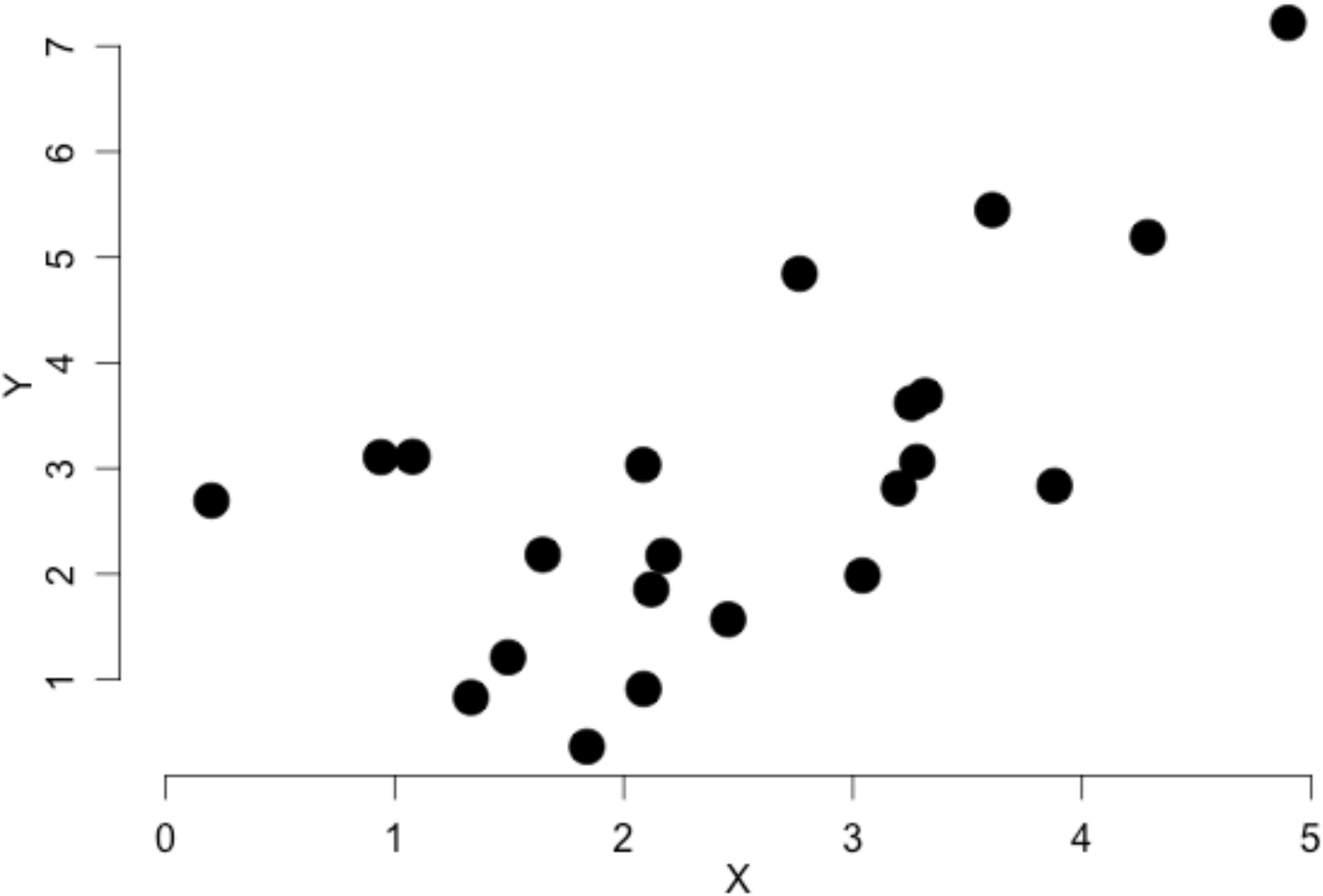
Which fits best? Model A or Model B?

In our case “Fit” is our friend, the likelihood function.

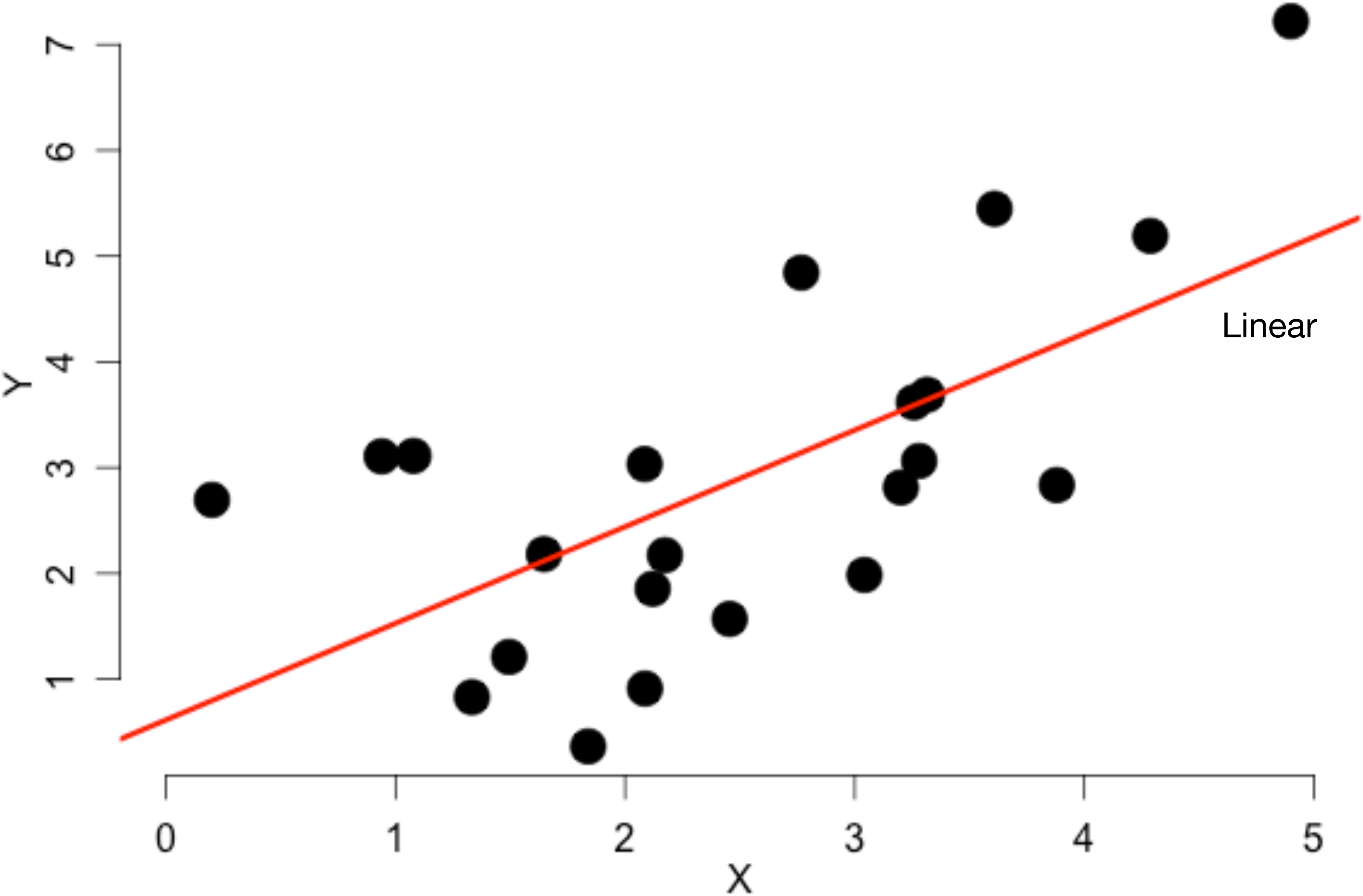
We calculate the **best fit** for two different models.

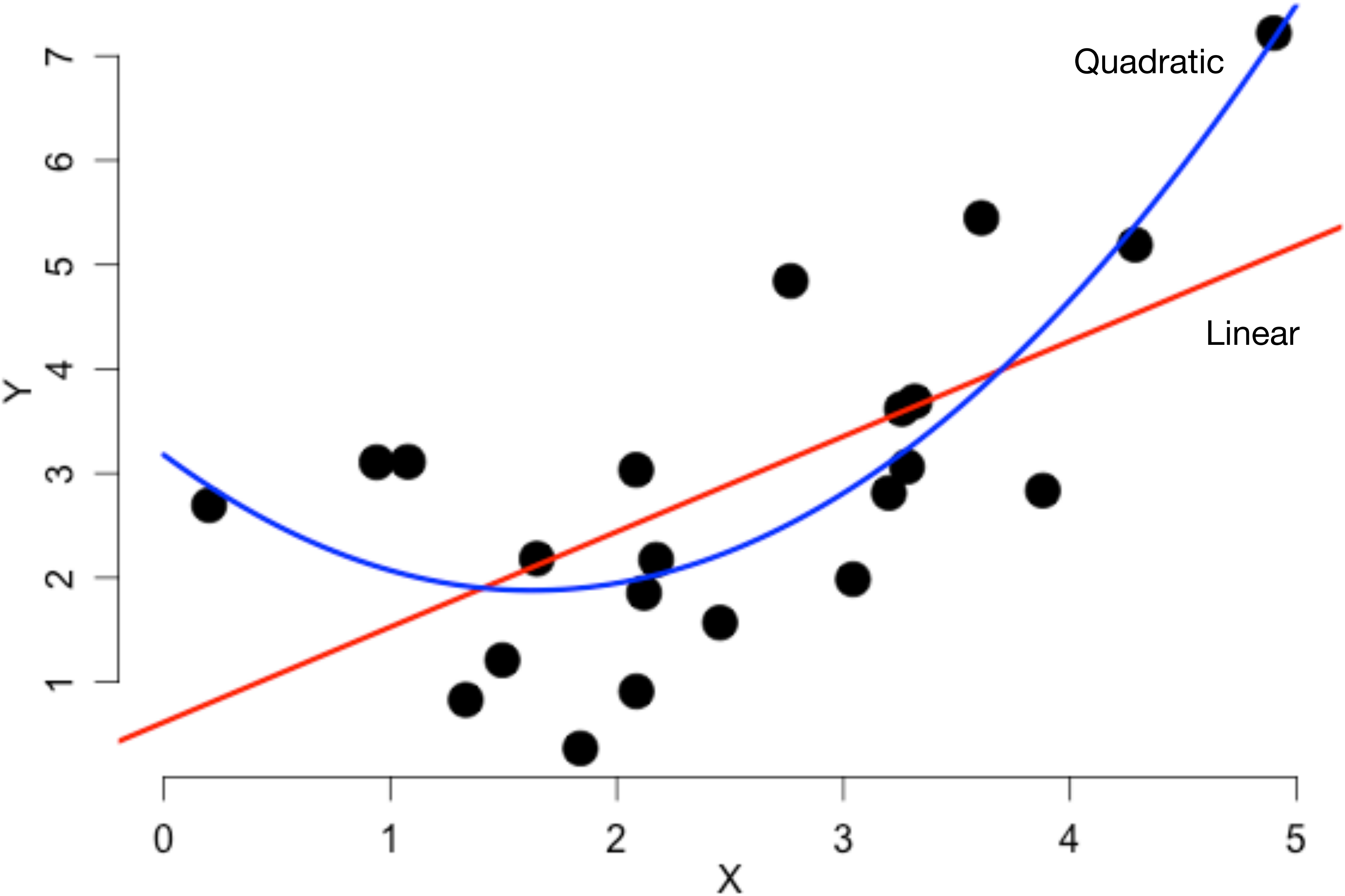
If the more complex model fits **much** better, than we choose it.

If it only fits a little better, we keep the simpler model.

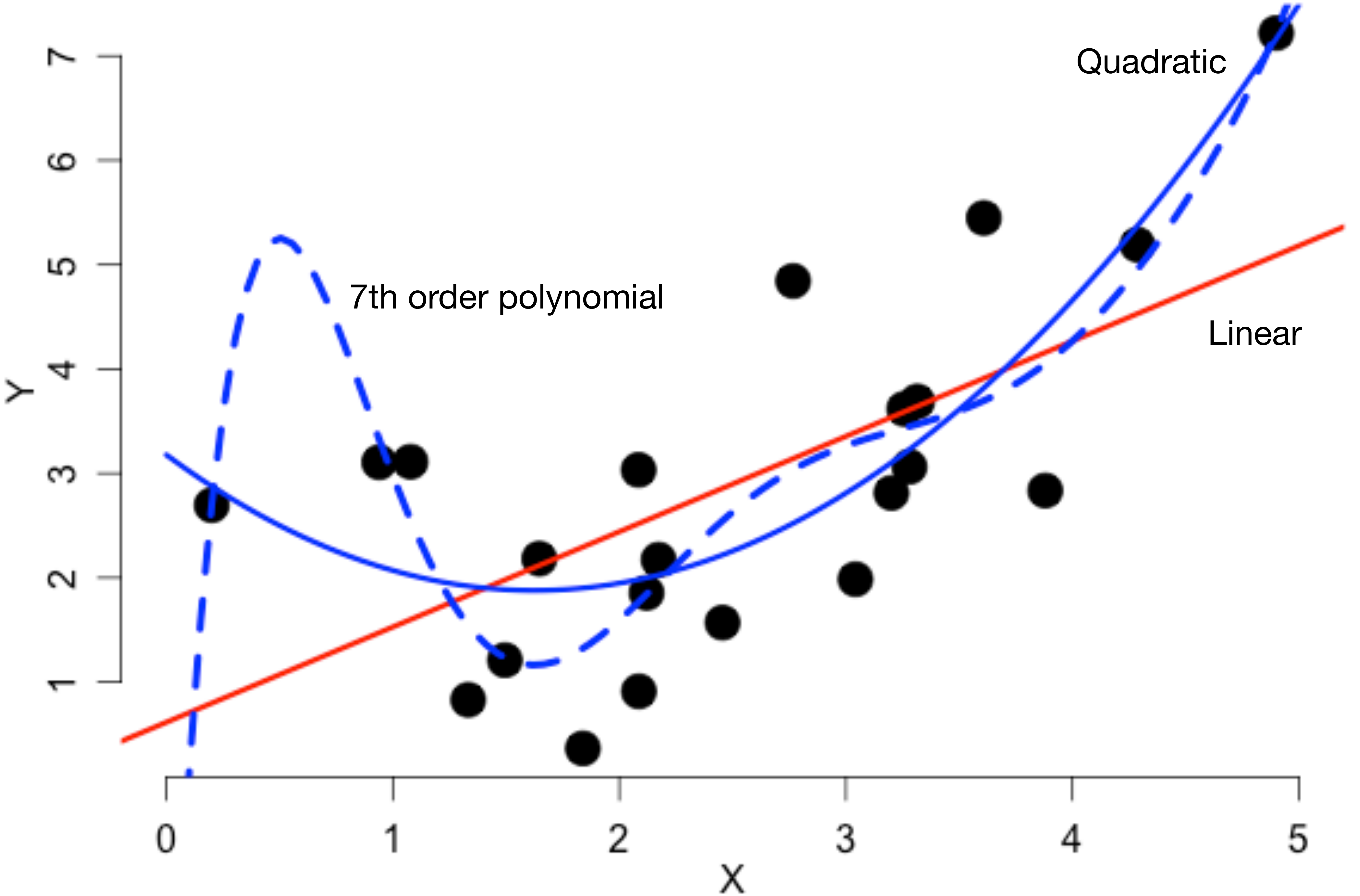


Y
7
6
5
4
3
2
1
0
1
2
3
4
5
X


Linear
X
Y
0
1
2
3
4
5
1
2
3
4
5
6
7



Quadratic
Linear
Y
X
0
1
2
3
4
5
1
2
3
4
5
6
7


Quadratic
7th order polynomial
Linear
Y
X
0
1
2
3
4
5
1
2
3
4
5
6
7

Our overall Log Likelihood for OLS Regression

$$\ell(\beta, \sigma^2; X, y) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i\beta)^2 - \frac{n}{2}\log(2\pi\sigma^2)$$

The Likelihood Ratio Statistic:

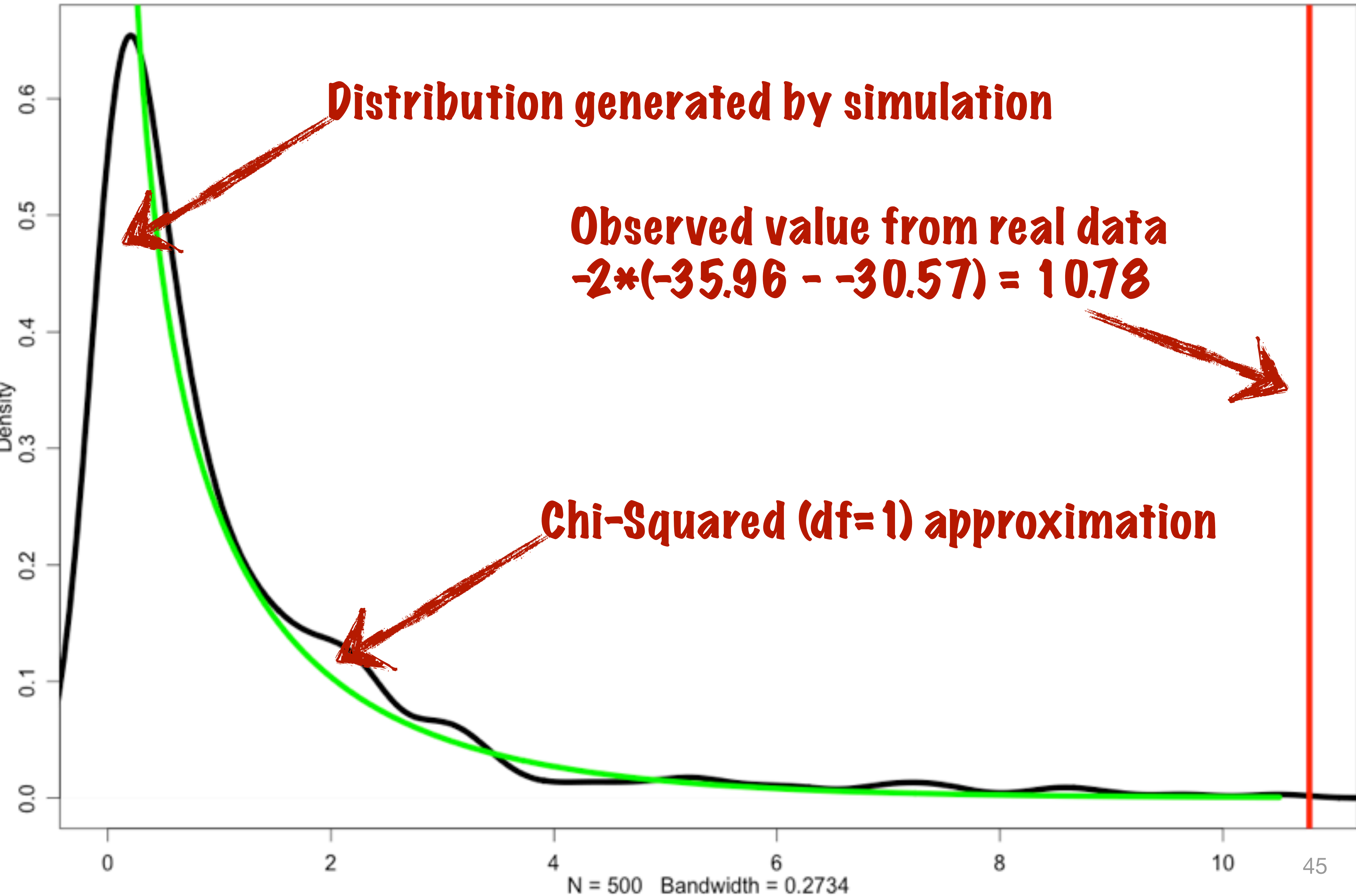
$$\begin{aligned} D &= -2\log\frac{L(small)}{L(big)} \\ &= -2\left(\ell(\widehat{\beta}_{small}; X_{small}, y) - \ell(\widehat{\beta}_{big}; X_{big}, y)\right) \end{aligned}$$

The Log-Likelihoods

```
> logLik( l1 )
'log Lik.' -35.95831 (df=3)

> logLik( l2 )
'log Lik.' -30.57043 (df=4)

> logLik( l7 )
'log Lik.' -27.93358 (df=9)

> -2*(-35.96 - -30.57)
[1] 10.78
```

Bigger models always fit better. The question is **do they fit more better than you would believe due to random chance**?

Distribution of the Test Statistic



The Distribution of the LR Test Statistic

For our test statistic

$$D = -2\left(\ell(\widehat{\beta}_{small}; X_{small}, y) - \ell(\widehat{\beta}_{big}; X_{big}, y)\right)$$

D is always greater than 0.

We have under the null (i.e., where the small model is true) that (approximately):

$$D \sim \chi^2_m$$

Where *m* is the degrees-of-freedom, as in the number of new parameters we need to estimate in the big model.

Likelihood Ratio Tests
(on our polynomial regression from before)

```
> library( lmtest )
> lrtest( l1, l2, l7 )
Likelihood ratio test

Model 1: Y ~ X
Model 2: Y ~ X + X2
Model 3: Y ~ X + X2 + X3 + X4 + X5 + X6 + X7
  #Df  LogLik Df   Chisq Pr(>Chisq)
1   3 -35.958
2   4 -30.570  1 10.7758   0.001028 **
3   9 -27.934  5  5.2737   0.383398
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

We see in the above:

1) Our 2nd model is better than our 1st

2) Our 3rd model is not significantly better than our second.

How to conduct the likelihood ratio test

1. Compute $2\Delta LL = \log(L_1) - \log(L_0)$
   This will always be negative, or at least non-positive.
2. Compute Δp = # parameters you just constrained
   This is equal to # parameters in full model - # parameters in reduced model.
3. Compare $-2\Delta LL$ to a χ^2 distribution with Δp degrees of freedom
   The chance of getting this number or larger is the *p*-value.

(Δ is the Greek letter "delta", and is typically used to denote differences or changes.)

The full and constrained model for our example

Fit one model where γ_{10} is freely estimated

```
mod <- lmer(mathach ~ 1 + ses + sector + (1 + ses|id),
     data = dat,
    REML = FALSE)
```

Be sure to use MLE, not REML

Fit a null model where is constrained to be 0

```
mod_null <- lmer(mathach ~ 1 + sector + (1 + ses|id),
    data = dat,
    REML = FALSE)
```

The Full Model

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Our Null

$$H_0 : \gamma_{10} = 0$$

In our model

Extract the log-likelihoods …

```
> ll1 <- logLik(mod)
> ll1
'log Lik.' -23299 (df=7)
> ll0 <- logLik(mod_null)
> ll0
'log Lik.' -23401 (df=6)
> delta_ll <- ll1 - ll0
> 2 * delta_ll
'log Lik.' 205 (df=7)
```

… and the number of parameters. There will be an easier way to do this, but note the logLik does give us number of parameters (see above) so:

```
> delta_p <- 7 - 6
```

In our model

So compare 2ΔLL to a χ^2 distribution.

```
> pchisq(2*as.numeric(delta_ll), delta_p,
                          lower.tail = FALSE)
[1] 1.46e-46
```

Note: the as.numeric is to tell R our ratio is just a number, not an example of the R "class" logLik. This as.numeric is optional; you will get the right p-value but the printout will look weird if you don't use it.

Making things a little simpler

Okay, okay, so, R can conduct that test for you.

We'll use the **anova** function, which also handles log-likelihood tests

Pro-tip: if you mistakenly use REML to fit the model, ANOVA will refit using ML before testing.

```
> anova( mod, mod_null )
mod_null: mathach ~ 1 + sector + (ses | id)
mod: mathach ~ 1 + ses + sector + (ses | id)
         Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
mod_null  6 46815 46856 -23401    46803
mod       7 46611 46660 -23299    46597   205      1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multiparameter Tests

Example: Does sector relate at all to math achievement?

$$Y_i = \beta_{0j[i]} + \beta_{1j[i]} ses_i + \epsilon_i$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

Our null hypothesis:

$$H_0 : \gamma_{01} = \gamma_{11} = 0$$

We want to test that sector has nothing to do with achievement or the achievement-ses relationship (i.e., sector can be dropped).

Two Options for Multiparameter Tests

General Linear Hypothesis Tests (testing *contrasts*)

- ★ Requires identical variance-covariance specification
- ★ Can be used for individual random effects as well
- ★ Generally superior for fixed effects
- ★ Some properties bad, can avoid problems with simulation.

Likelihood Ratio Tests

- ★ Need Maximum Likelihood models (not restricted maximum likelihood)
- ★ Not as flexible (e.g., can't test the difference of two coefficients)
- ★ Good for testing whether things should be included in a model.

Option 1: Contrasts

Multiparameter tests of fixed effects are often called *contrasts*.

Look at the 'contest()' method in lmerTest

These tests are rarely needed.

"contrast" comes from old ANOVA and randomized experiment literature.

Option 2: Likelihood ratio test!

Same idea as we just talked about...

Maximum Likelihood on Multiple Fixed Effects

```
> M1 = lmer( mathach ~ 1 + ses * sector + (1 + ses|id),
         data=dat, REML=FALSE )

> M0 = lmer( mathach ~ 1 + ses + (1 + ses|id),
         data=dat, REML=FALSE )
```

This says "Do not use restricted maximum likelihood."

```
> lrtest( M0, M1 )
Likelihood ratio test

Model 1: mathach ~ 1 + ses + (1 + ses | id)
Model 2: mathach ~ 1 + ses * sector + (1 + ses | id)
  #Df LogLik Df Chisq Pr(>Chisq)
1   6 -23318
2   8 -23282  2 73.29  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Note the Df=2 since we are testing two parameters here!

Big takeaways on testing fixed effects

Big takeaways on testing fixed effects

You have a couple of tools to conduct significance tests for fixed effects.

They won't give identical results, but should tend to give substantively the same answer

Likelihood ratio tests (LRTs) are a good, general tool.

You certainly sound cooler when you talk about them.

I mean, *"Wald test"*? That doesn't sound so great.

Inference for variance parameters

$$y_i = \alpha + \alpha_{j[i]} + (\beta + \beta_{j[i]})x_i + \epsilon_i$$

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

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_\alpha^2 & \sigma_\alpha \sigma_\beta \rho \\ \sigma_\alpha \sigma_\beta \rho & \sigma_\beta^2 \end{pmatrix}\right]$$

Main tool: the Likelihood Ratio Test (again)

Let's test whether we need a random slope:

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

$$H_0 :???$$

What are our nested models?

MLM Models and Likelihood Ratios

```
> lrtest( M1, M2 )
Likelihood ratio test

Model 1: mathach ~ 1 + ses + sector + (1 | id)
Model 2: mathach ~ 1 + ses + sector + (1 + ses | id)
  #Df LogLik Df  Chisq Pr(>Chisq)
1   5 -23306
2   7 -23301  2 9.2963   0.009579 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 
0.1 ' ' 1
```

Why 2 degrees of freedom here?

We have reason to believe random slopes are important!

A test with level-2 predictors included

```
> lrtest( M1, M2 )
Likelihood ratio test

Model 1: mathach ~ 1 + ses * sector + (1 | id)
Model 2: mathach ~ 1 + ses * sector + (1 + ses | id)
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 -23287
2   8 -23285  2 5.0521    0.07997 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

Now the model is only marginally better.

What is going on?

Aside: the Boundary Issue for testing variance parameters

We expect an unbiased estimator of a parameter to be below 50% of the time and above 50% of the time.

So under the null, we would expect negative variance 50% of the time.

Our null sampling distribution is 50% 0 and 50% our usual chi-squared.

This means we expect the distribution of our likelihood ratio (the test statistic) for testing the q+1th random effect to be:

$$0.5\chi^2(q) + 0.5\chi^2(q+1)$$

Upshot: for random intercept models, divide *p*-values by 2

In general the naïve approach is *conservative* meaning that the *p*-value will be too high. So if you reject using the naive approach, you reject using the corrected approach. In my mind, this means use the naïve approach without fear.

RH&S 2.6.2 (pg 88-89)
Also see RH&S 3.6.3, and 4.6

Recap & reflect time

Recap

Check-In
http://cs179.org/lec26

There are several tools for inference and uncertainty estimation

These tools are generally exchangable, but two go-to tools for almost everything are:

- ★ Profile confidence intervals. It asks “what is a reasonable range for what this parameter might actually be, given the other parameter estimates?”
- ★ The Likelihood Ratio Test. It asks “is my more flexible model much, much better than my less flexible one?”

Inference for fixed effects is easy and robust

Inference for variance components is shaky but ok

Inference for individual random effects is poor and should be avoided.

Appendix

Some further extensions and notes

Inference for **individual** random effects

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \beta_2 fem_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2)$$

$$\beta_0 j = \gamma_{00} + \gamma_{01} sector_j + u_{0j}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} sector_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right]$$

Testing a single random effect

Our null:

$$H_0 : \beta_{qj} = 0 \text{ for some specific } j$$

Our test statistic:

$$z = \frac{\beta^*_{qj}}{\sqrt{V^*_{qqj}}}$$

We need normal approximation here. Under the null, *z* is distributed N(0, 1)

Do Confidence Intervals Instead

Warning:
These tests have bad properties

Alternatively, just analyze the group of interest.

(Which will likely be too low power for use due to radically reduced amounts of data.)

Standard Errors and MLE

We're not going to show this in detail, but we can get standard errors based on the curvature of the likelihood function evaluated at the MLE estimate

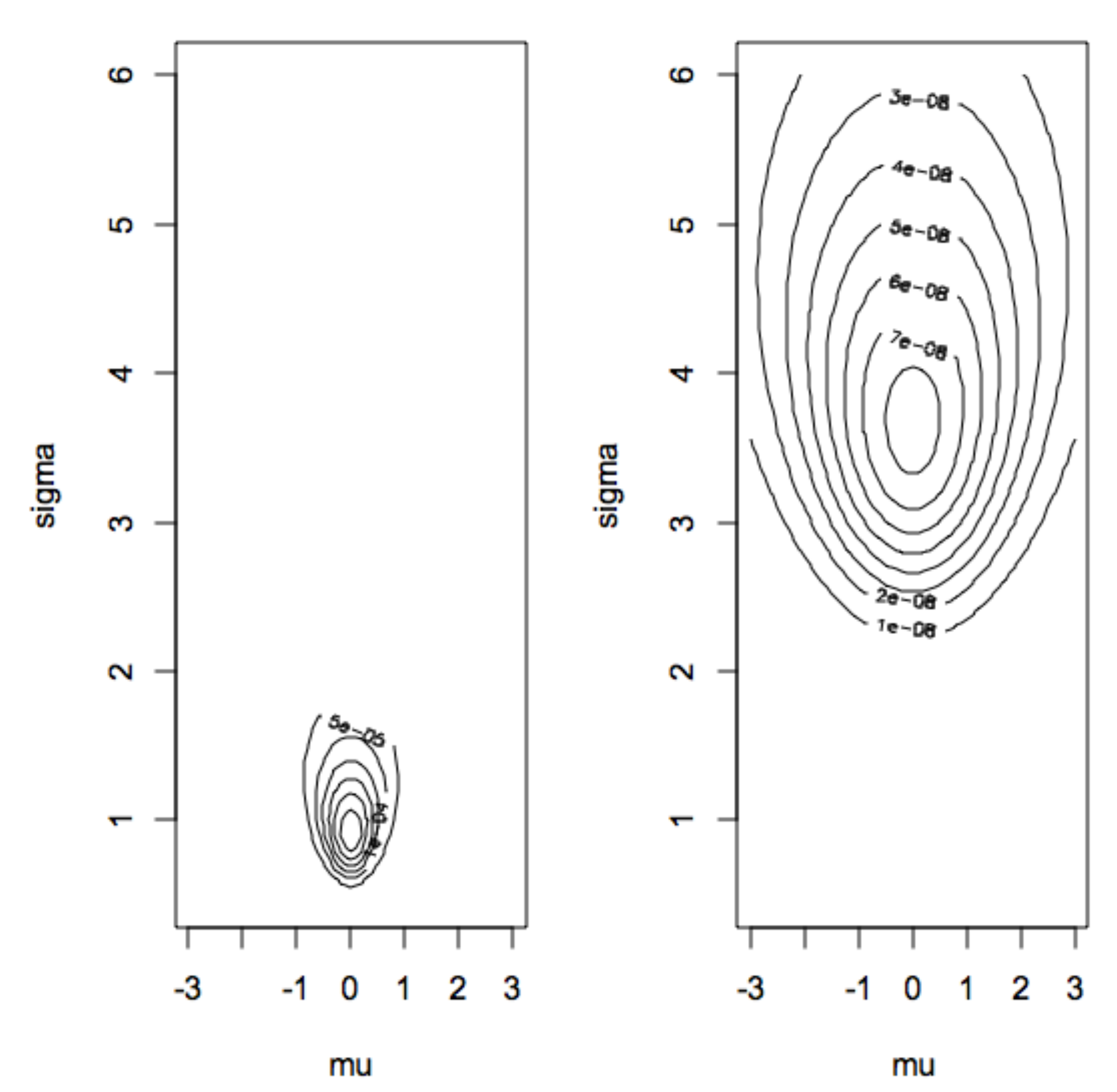
If there's **lots of curvature** (the steepness of the hill is changing quickly near the top), **our estimates are more precise**.

- ★ Only a small set of parameter values have good values, so we can easily locate what parameters are best.

If there's **not much curvature** (the hill is really flat near the top), our estimates are **less precise**.

- ★ Lots of parameter values are all nearly as good as the top and we can't distinguish between then.

A steep likelihood and a flat likelihood



Profile Confidence Intervals

Here's another approach.

We'll focus on σ_0, but we can do the same thing for each parameter

- ★ Start by proposing a value of σ_0^{hyp}
- ★ Find the values of the other parameters which maximize the likelihood subject to this constraint $(\hat{\mu}_\alpha, \hat{\mu}_\beta, \hat{\sigma}_1, \hat{\rho}, \hat{\sigma}_\varepsilon)$.
- ★ Now do the same likelihood ratio testing as proposed before at this particular point.
- ★ Repeat until we find the edges of the confidence interval.

Vastly reduces the complexity of the task.

Profile Confidence Intervals, Continued

Say we want the profile likelihood for β_0

For each value β_0 can take (e.g., $\beta_0 = 1.21$)

- ★ Find the maximum likelihood estimates for the other parameters (β_1 , σ, etc) given $\beta_0 = 1.21$
- ★ Finally consider whether $\beta_0 = 1.21$ is plausible by seeing if the overall model fits the data well.
- ★ If it is, put it in the confidence interval. If it isn't, reject it.

(This is called "test inversion." We make a confidence interval out of all values that are plausible given the data.)