STA303/1002 - Methods of Data Analysis II
(Week 11 lecture note)

Wei (Becky) Lin

Mar 21/23, 2017

UNIVERSITY OF TORONTO

Topics learned last week

- Poisson regression model
  - Deivance

$$D = 2\sum_{i=1}^{n}\left(y_i \log\frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i)\right)$$

  - Inference about individual β_j
  - Goodness of fit tests: (1) LRT using deviance (2) Pearson χ^2 test.
  - 3 types of residuals: response/Pearson/Diviance residual.
  - Residual plots.
  - Overdispersion: $\hat{\Phi} = X^2/(n-p)$ (default in R) or $\hat{\phi} = D/(n-p)$ and $\hat{\Phi} >> 1$
  - Two ways to get summary output adjusted by overdispersion
- Log-linear model for 2-way contingency table
  - To test X and Y (row and column variables) are independent
    - Fisher's Exact Test (assume row total is fixed.)
    - Pearson χ^2 test
    - Deviance test (LRT) test
    - Testing the interaction term is significant or not
    - Using logit model and test whether the slope β_1 is zero or not.

*

Topics learned last week

- Log-linear model for 3-way contingency table
- Y_{ijk} count in cell (i,j,k), $Y_{ijk} \sim Pois(\lambda_{ijk})$
- Conditional on $n = \sum_{i,j,k} n_{ijk}$

$$Y_{ijk} | n \sim Multinom(\pi_{ijk})$$

- $\mu_{ijk} = \log E(Y_{ijk})$-reference cell
- $\alpha_i, \beta_j, \gamma_k$: deviations of $\log E(Y_{ijk})$ from reference cell, $\alpha_1 = \beta_1 = \gamma_1 = 0$
- Residual df=IJK- # of parameters.

Model	$\log E\{Y_{ijk}\} =$
Mut. Indep	$\mu + \alpha_i + \beta_j + \gamma_k$
Joint Indep.	$\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
Cond. Indep.	$\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
Unif. Assoc.	$\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
Saturated	$\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$

Model	$\pi_{ijk} =$	$\hat{\lambda}_{ijk} =$ $n_{ijk} = \hat{Y}_{ijk} = \hat{E}_{ijk}$
Mut. Indep	$\pi_i \pi_j \pi_k$	$n_{i++}n_{+j+}n_{++k}/n^2$
Joint Indep.	$\pi_{ij}\pi_k$	$n_{ij+}n_{++k}/n$
Cond. Indep.	$\pi_{ik}\pi_{jk}/\pi_k$	$n_{i+k}n_{+jk}/n_{++k}$
Unif. Assoc.	$\pi_{ij}\pi_{ik}\pi_{jk}$	Iterative
Saturated	π_{ijk}	n_{ijk}

$\lambda_{ijk} = n\pi_{ijk}$

Learning Objectives This Week

Week 11 and Week 12 notes are based on KNNL ch25.

- Introduction to LMM (Linear Mixed Models, or so called Linear Models with Random Effect)
- One-way Random Effects model
  - ANOVA based estimation
  - ML estimation
  - REML estimation
- Two-way Random Effects model
  - ANOVA based estimation
  - ML estimation
  - REML estimation

Introduction to LMM

So far: independent response variable, but often

- Clustered Data
  - Response is measured for each subject
  - each subject belongs to a group of subjects (cluster)
  - Examples
    - STA scores of student grouped by classrooms (class room forms cluster)
    - birth weight of rats grouped by litter (litter forms cluster)
- Longitudinal Data
  - response is measured at several time points
  - number of time points is not too large (in contrast to time series)
  - Example: sales of a product at each month in a year (12 measurement)

Fixed and Random Factors/Effects

How can we extend the linear model to allow for such dependent data structures?

Fixed factor = **qualitative** covariate (e.g. gender, agegroup)

Fixed effect = **quantitative** covariate (e.g. age)

Random factor = **qualitative** variable whose levels are randomly sampled from a population of levels being studied

Random effect = **quantitative** variable whose values are randomly sampled from a population of values being studied

Random factor example: 20 supermarkets were selected and their number of cashiers were reported. The observed levels of random factor "number of cashier"

- 10 supermarkets with 2 cashier
- 3 supermarkets with 1 cashier
- 7 supermarkets with 6 cashier

Random effect example: 20 supermarkets were selected and their size reported. These size values are random samples from the population of size values of all supermarkets.

One-way Random Effects Model

Example: Rating data of job applicants

- Data is from KNNL p.1036
- Interested in studying the variability in the rating of job applicants
  - Variability among applicants
  - Variability among personnel officers.

Data:

- Y is the job applicant rating
- Factor: officer/interviewer (r=5)
- Interviewers selected **at random** from population of personnel officers
- 20 applicants randomly and equally assigned (n=4) to officers.

Balanced design

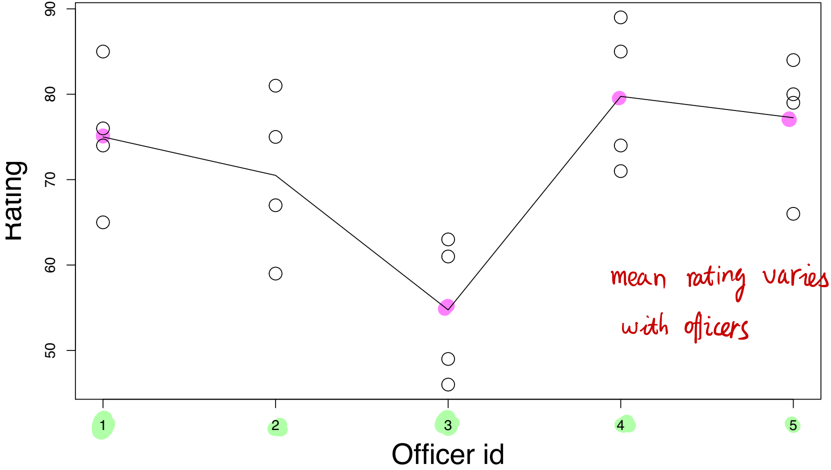
```
job = read.table("CH25TA01.txt", sep="",as.is=T,header=F)
colnames(job) = c("rating","officer", "replicate")
job$officer = as.factor(job$officer)
head(job)
```

```
##   rating officer replicate
## 1     76       1         1
## 2     65       1         2
## 3     85       1         3
## 4     74       1         4
## 5     59       2         1
## 6     75       2         2
```

Example: Rating data of job applicants

Mean plot

```
plot(job$officer,job$rating, xlab="Officer id", ylab="Rating",cex=2,cex.lab=2)
lines(1:5, with(job,tapply(rating, officer,mean)))
```



Data for One-way Random Effects model

- Y is the response variable
- Factor with levels i=1,2,. . . r
- Y_{ij} is j-th observation from cell i
- Consider $j = 1, \ldots, n$

level	Observations
1	$y_{11}, \quad y_{12}, \quad \ldots \quad y_{1n}$
2	$y_{21}, \quad y_{22}, \quad \ldots y_{2n}$
$\vdots$	$\vdots$
r	$y_{r1}, \quad y_{r2}, \quad \ldots \quad y_{rn}$

① belongs to same group, correlated

② different groups. data are independent.

Random Effect model

- Cell means model

R.V., not a parameter

$$Y_{ij} = \mu_i + \epsilon_{ij} \Rightarrow \begin{cases} E(Y_{ij}) = E(\mu_i) + E(\epsilon_{ij}) = \mu \\ V(Y_{ij}) = \sigma_\mu^2 + \sigma^2 \\ Y_{ij} \sim N(\mu, \sigma^2 + \sigma_\mu^2) \end{cases}$$

  - $\mu_i \sim_{iid} N(\mu, \sigma_\mu^2), \epsilon_{ij} \sim_{iid} N(0, \sigma^2)$
  - μ_i and ϵ_{ij} are independent.

- Factor Effect model

R.V. not a parameter.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

  - $\tau_i \sim_{iid} N(0, \sigma_\mu^2), \epsilon_{ij} \sim_{iid} N(0, \sigma^2)$
  - τ_i and ϵ_{ij} are independent.

- Called Model II in KNNL book.

Equivalent Specification in Terms of Conditional Distribution

$$* \begin{cases} E(Y) = E(E(Y|X)) \\ V(Y) = E(V(Y|X)) + V(E(Y|X)) \end{cases}$$

- Cell mean model

$$\mu_i \sim_{iid} N(\mu, \sigma_\mu^2)$$

$$Y_{ij}|\mu_i \sim_{ind} N(\mu_i, \sigma^2)$$

$$E(Y_{ij}) = E(E(Y_{ij}|\mu_i)) = E(\mu_i) = \mu$$

$$V(Y_{ij}) = E(V(Y_{ij}|\mu_i)) + V(E(Y_{ij}|\mu_i)) = E(\sigma^2) + V(\mu_i) = \sigma^2 + \sigma_\mu^2$$

- Factor effects model

$$\tau_i \sim_{iid} N(0, \sigma_\mu^2)$$

$$Y_{ij}|\tau_i \sim_{ind} N(\mu + \tau_i, \sigma^2)$$

$$\begin{cases} E(Y_{ij}) = \mu \\ V(Y_{ij}) = \sigma^2 + \sigma_\mu^2 \end{cases}$$

- Similarity to Bayesian methods
  - parameters μ_i have a distribution
- Difference from Bayesian methods
  - $\mu, \sigma^2, \sigma_\mu^2$ do not have a distribution.

Implications of the Random Effects Model

- There are TWO variance parameters.
- Cell means (μ_i) are random variables, not parameters.
- $Y_{ij} \sim N(\mu, \sigma_\mu^2 + \sigma^2)$
- The observations are not independent:

$$\boxed{cov(Y_{ij}, Y_{i'j'}) = 0, \quad cov(Y_{ij}, Y_{ij'}) = \sigma_\mu^2}$$

- e.g. if r=2 and n=2

var(Y11) cov(Y11, Y12) r=1 group

$$Var \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \begin{bmatrix} \sigma_\mu^2 + \sigma^2 & \sigma_\mu^2 & 0 & 0 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma_\mu^2 + \sigma^2 & \sigma_\mu^2 \\ 0 & 0 & \sigma_\mu^2 & \sigma_\mu^2 + \sigma^2 \end{bmatrix}$$

cov(Y22, Y11) = 0

different group

r=2 group

ANOVA Table

$N(0,\sigma_\mu^2)$ $N(0,\sigma^2)$

under $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

Source	d.f.	SS	EMS
Trmt	r-1	$\sum_i^r n(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$\sigma^2 + n\sigma_\mu^2$
Error	nr-r	$\sum_i^r \sum_j^n (\bar{Y}_{ij} - \bar{Y}_{i.})^2$	σ^2
Total	nr -1		

If True, $\tau_i = 0 \Rightarrow Y_{ij} = \mu + \varepsilon_{ij} \Rightarrow \mu_i = \mu \ \forall i$

- $H_0 : \sigma_\mu^2 = 0, \quad vs \quad H_a : \sigma_\mu^2 > 0$ $\Leftrightarrow$ $H_0: \mu_1 = \mu_2 = \cdots = \mu_r$, $H_a: \mu_i \neq \mu_j$ some i,j
- $F_0 = \frac{MS_{trmt}}{MSE} \sim_{H_0} F_{r-1,nr-r}$
- Conclusion pertains to entire population.
- Reject H_0:
  - The expected ratings of the population of officers has a non-zero variance. $\sigma_\mu^2 \neq 0$
  - The company need to improve consistency between the interviewers.

Inference About $E(Y_{ij})$

$Var(\bar{Y}_{..}) = var(\frac{1}{nr}\sum_{i=1}^{r}\sum_{j=1}^{n} Y_{ij})$

$= Var(\frac{1}{nr}\sum_{i=1}^{r}(\mu_i + \varepsilon_{i1} + \cdots + \varepsilon_{in}))$

$= \frac{1}{nr}\sum_{i=1}^{r}(V(\mu_i) + V(\varepsilon_{i1}) + \cdots V(\varepsilon_{in}))$

$= \frac{1}{nr}(\sigma_\mu^2 + n\sigma^2)$

It's an unknown constant, i.e. a model parameter.

Inference about $\mu = E(Y_{ij})$

- $E(Y_{ij}) = \mu$
- $\widehat{E(Y_{ij})} = \bar{Y}_{..}$
- $Var(\bar{Y}_{..}) = \frac{\sigma_\mu^2}{r} + \frac{\sigma^2}{rn} = \frac{n\sigma_\mu^2 + \sigma^2}{rn}$
- $\widehat{Var(\bar{Y}_{..})} = \frac{MS_{trmt}}{rn}$
- Testing:

$$\boxed{T = \frac{\bar{Y}_{..} - \mu}{\sqrt{MS_{trmt}/rn}} \sim_{H_0} t_{r-1}}$$

$\Rightarrow df(SS_{trmt}) = r-1$

$$T = \frac{\hat{\theta} - \theta}{\sqrt{\widehat{Var}(\hat{\theta})}}$$

- $\theta = \mu$
- $\hat{\theta} = \bar{Y}_{..}$
- $var(\hat{\theta}) = var(\bar{Y}_{..}) = \frac{n\sigma_\mu^2 + \sigma^2}{nr}$

$\Rightarrow \widehat{var(\hat{\theta})} = \frac{n\hat{\sigma}_\mu^2 + \hat{\sigma}^2}{nr} = \frac{MS_{trmt}}{nr}$

- $(1 - \alpha)$ CI for μ

$$\bar{Y}_{..} \pm t_{1-\alpha/2, r-1} \sqrt{MS_{trmt}/rn}$$

$\hat{\theta} \pm$ Critical value $\cdot \widehat{SE}(\hat{\theta})$

Intraclass Correlation Coefficient (ICC)

- Percentage of total variation due to factor

$$\boxed{\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} = \frac{\sigma_\mu^2}{\sigma_Y^2}}$$

- Correlation between two observations with the same i (e,g. between two evaluations of a same officer)

$$\boxed{\rho_{IC} = \frac{cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})Var(Y_{ik})}} = \frac{\sigma_\mu^2}{\sigma_Y^2}}$$

σ_μ^2 rel. smaller

- Smaller intraclass correlation implies little variation among officers.
- Larger intraclass correlation implies little variation among applicants.

$\rho_{IC} = \frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}$ σ^2 relatively larger $\to \rho_{IC} \uparrow$

Confidence Interval for ρ_{IC}

CI for $\rho_{IC} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$

- MS(trmt) and MSE are independent r.v.

$$\boxed{\frac{MS_{trmt}}{n\sigma_\mu^2 + \sigma^2} \Big/ \frac{MSE}{\sigma^2} \sim F_{r-1,nr-r}}$$

$$P\left\{F_{\alpha/2,r-1,nr-r} \leq \frac{MS_{trmt}}{MSE}\frac{\sigma^2}{n\sigma_\mu^2 + \sigma^2} \leq F_{1-\alpha/2,r-1,nr-r}\right\} = 1 - \alpha$$

- Solve for $\rho_{IC} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$

$$\frac{L}{L+1} \leq \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} \leq \frac{U}{U+1}$$

where

$$L = \frac{1}{n}\left(\frac{MS_{trmt}}{MSE\ F_{1-\alpha/2,r-1,nr-r}} - 1\right)$$

$$U = \frac{1}{n}\left(\frac{MS_{trmt}}{MSE\ F_{\alpha/2,r-1,nr-r}} - 1\right)$$

- If low limit of the CI is negative, set to zero.

Confidence Interval for σ^2

Confidence Interval for σ^2 ← ε_{ij}

- $\hat{\sigma}^2 = MSE$
- Known distribution of a function of MSE

$$\boxed{\frac{r(n-1)MSE}{\sigma^2} \sim \chi^2_{nr-r}}$$

$$P\{\chi^2_{\alpha/2} \leq \frac{r(n-1)MSE}{\sigma^2} \leq \chi^2_{1-\underset{2}{\alpha},nr-r}\} = 1-\alpha$$

- Solve for σ^2

$$\frac{r(n-1)MSE}{\chi_{1-\alpha/2,nr-r}} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2_{\alpha/2,nr-r}}$$

- Replace nr with n_T in unbalanced designs.

↳ (refer to KNNL)

Point Estimate for σ_μ^2

Point Estimate for σ_μ^2

- $E\{MS_{trt}\} = \sigma^2 + n\sigma_\mu^2$
- $E\{MSE\} = \sigma^2$

$\Rightarrow (\sigma^2 + n\sigma_\mu^2) - \sigma^2 = n\sigma^2$

- σ_μ^2

$$\sigma_\mu^2 = \frac{E(MS_{trt}) - E(MSE)}{n} = \frac{1}{n}E(MS_{trt}) - \frac{1}{n}E(MSE)$$

- $\hat{\sigma}_\mu^2$

$$\boxed{\hat{\sigma}_\mu^2 = \frac{MS_{trt} - MSE}{n} = \frac{1}{n}MS_{trt} - \frac{1}{n}MSE}$$

- $\hat{\sigma}_\mu^2$ is estimated by a linear combination of independent MS
- Can adjust denominator in unbalanced experiments.
- σ_μ^2 can be negative.

Confidence Interval for σ_μ^2

- $\sigma_\mu^2 = c_1 E(MS_1) + \ldots + c_h E(MS_h) = L$
- $\hat{\sigma}_\mu^2 = c_1 MS_1 + \ldots + c_h MS_h = \hat{L}$
- $\frac{df\ \hat{L}}{L} \sim_{approx} \chi^2_{df}$

$$P\{\chi^2_{\alpha/2} \leq \frac{df\ \hat{L}}{L} \leq \chi^2_{1-\alpha/2,df}\} =_{approx}\ 1-\alpha$$

- Solve for L

$$\frac{df\ \hat{L}}{\chi^2_{1-\alpha/2,df}} \leq L \leq \frac{df\ \hat{L}}{\chi^2_{\alpha/2,df}}$$

- Satterwaite approximation of df

$$df = \frac{(c_1 MS_1 + \ldots + c_h MS_h)^2}{\frac{(c_1 MS_1)^2}{df_1} + \ldots + \frac{(c_h MS_h)^2}{df_h}}$$

- use the nearest integer for the df

Rating data of Job applicants: ANOVA-based estimation

- Use **aov** in R; could also use **lm** or **glm**

```
job$officerFac <- factor(job$officer)
# ========== Fixed-effects ANOVA =========
mod1 <- aov(rating~ officerFac,data=job)
summary(mod1)
```

$H_0: \sigma_\mu^2 = 0$

$H_a: \sigma_\mu^2 > 0$

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## officerFac    4   1580   394.9   5.389 0.0068 **
## Residuals    15   1099    73.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MStrmt

MSE

```
coef(mod1)
```

```
## (Intercept) officerFac2 officerFac3 officerFac4 officerFac5
##       75.00       -4.50      -20.25        4.75        2.25
```

```
# === Calculate estimate of sigma2_mu =====
table(job$officerFac)
```

```
##
## 1 2 3 4 5
## 4 4 4 4 4
```

← check $n_1 = n_2 = \cdots = n_r$: balanced design

```
( sigma2mu.hat = (394.93-73.28)/4 )
```

← $\hat{\sigma}_\mu^2 = \dfrac{MStrmt - MSE}{n} = 80.4125$

```
## [1] 80.4125
```

Pros and Cons of ANOVA-based Inference

- Advantages
  - Explicit formulae
  - Clear insight into the mechanism
- Disadvantages
  - $\hat{\sigma}^2_\mu$ can be negative
  - For unbalanced designs and in presence of multiple factors, ANOVA decomposition is not unique (depends on the order of the factors), and SS are not orthogonal. Therefor the inference does not hold
  - Formulae become more complex with more factors.
- Alternative solution
  - Maximum likelihood (ML) or restricted maximum likelihood (REML)

Maximum Likelihood Estimation

var-cov for group 1 Y_{1j}

$$V = \begin{pmatrix} \square & & \\ & \square & 0 \\ 0 & & \square \end{pmatrix}$$

- Define $Y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ and V_i the i-th block of the variance matrix V.
- Assumptions: $Y_i \sim MVN(\mu 1_{n_i}, V_i)$
- Likelihood:

$$L = \prod_i^r (2\pi)^{-n_i/2} |V_i|^{-1/2} \exp\{-\frac{1}{2}(y_i - \mu 1_{n_i})' V_i^{-1} (y_i - 1_{n_i}\}$$

- Solutions maximizing log-likelihood in balanced case

MLE ⇒
  - $\dot{\mu} = \bar{Y}_{..}$ ✓
  - $\dot{\sigma}^2 = MSE$ ✓
  - $\dot{\sigma}^2_\mu = \frac{1}{n}[(1 - 1/r)MS_{trmt} - MSE]$ different and it's a biased estimator.

ANOVA-based: $\hat{\sigma}_\mu^2 = \frac{MS_{trmt} - MSE}{n}$

Problems with MLE

so $\hat{\sigma}_\mu^{2\,(MLE)}$ is biased.

- $\dot{\sigma}_\mu^2$ is biased towards smaller values

$$E(\dot{\sigma}_\mu^2) = \frac{1}{n}\left[(1-1/r)\underbrace{E(MS_{trmt})}_{=\sigma^2+n\sigma_\mu^2} - \underbrace{E(MSE)}_{=\sigma^2}\right] = (1-1/r)\sigma_\mu^2 - \sigma^2/rn \neq \sigma_\mu^2$$

- $\dot{\sigma}_\mu^2$ can still be negative
- Solutions in the restricted parameter space

$$\hat{\sigma}_\mu^{2(MLE)} = \begin{cases} \dot{\sigma}_\mu^2 & \text{if } \dot{\sigma}_\mu^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\sigma}^{2(MLE)} = \begin{cases} \dot{\sigma}^2 & \text{if } \dot{\sigma}_\mu^2 > 0 \\ \frac{SS_{tot}}{nr} & \text{otherwise} \end{cases}$$

Rating data of Job Applicants: ML estimation

a unknown constant

rating $= \mu + \tau_i + \varepsilon_{ij}$

each officer introduces a random intercept.

- Use **lmer** in R, could also use **lme** (older)

random intercept

```
library(lme4)
mod2 <- lmer(rating~ 1+(1|officerFac),data=job,REML=FALSE)
```

group by officer

use MLE

- The data are grouped by "officer"
- The random effect is constant within each group

Rating data of Job Applicants: ML estimation

- Due to the bias, the ML-based estimate of variance is smaller than the ANOVA-based estimate
- This can result in too much optimism.

```
summary(mod2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: rating ~ 1 + (1 | officerFac)
##    Data: job
##
##      AIC      BIC   logLik deviance df.resid
##    156.0    158.9    -75.0    150.0       17
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4746 -0.8659  0.2453  0.5930  1.3054
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  officerFac (Intercept) 60.66    7.789
##  Residual               73.28    8.561
## Number of obs: 20, groups:  officerFac, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   71.450      3.975   17.98
```

$\hat{\mu}$

ANOVA: (Slide 21)

$\hat{\sigma}_u^2 = 80.4125$

$\hat{\sigma}^2 = 73.3$

MLE:

$\hat{\sigma}_u^{2(MLE)} = 60.66$

$\hat{\sigma}^{2(MLE)} = 73.28$

Restricted/ Residual ML (REML)

RE + ML = REML

- Apply ML to linear combinations of y, $K'y$
  - K selected s.t. the distribution of $K'y$ does not involves μ (or, more generally, any fixed effects)
  - Estimates of variance components are invariant to fixed effects
  - Implicitly takes into account the df for fixed effects.
- Simple example
  - suppose $Y_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$
  - Define $\bar{Y} = \sum_i^n Y_i/n, S_{yy} = \sum_i^n (Y_i - \bar{Y})^2$
  - MLE $\Rightarrow$ $\hat{\sigma}^2_{ML} = S_{yy}/n$
  - REML $\Rightarrow$ $\hat{\sigma}^2_{REML} = S_{yy}/(n-1)$, (unbiased)
- Estimate fixed effects as a second step

REML in one-way Random Effects ANOVA

- Factor effect model
$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$
  - The part of the likelihood that does not involve fixed effects is the part that does not involve μ
$$L(\mu, \sigma^2, \sigma_\mu^2 | Y) = L(\mu | \bar{Y}_{..}) \cdot \boxed{L(\sigma^2, \sigma_\mu^2 | SS_{trmt}, SSE)} \Rightarrow \hat{\sigma}_\mu^2$$
  - Use the second product in the likelihood as the likelihood for REML

↳ (but not a corresponding PDF, so not a genuine likelihood

- Solutions in balanced case:
  - $\boxed{\dot{\mu} = \bar{Y}_{..}, \dot{\sigma}^2 = MSE, \dot{\sigma}_\mu^2 = \frac{1}{n}[MS_{trmt} - MSE]}$
- Solutions maximizing log-likelihood in the restricted parameter space (same as ANOVA-based)

REML ⟺ ANOVA-based estimation

$$\hat{\sigma}_\mu^{2(REML)} = \begin{cases} \dot{\sigma}_\mu^2 & \text{if } \dot{\sigma}_\mu^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\sigma}^{2(REML)} = \begin{cases} \dot{\sigma}^2 & \text{if } \dot{\sigma}_\mu^2 > 0 \\ \frac{SS_{tot}}{nr} & \text{otherwise} \end{cases}$$

Rating data of Job Applicant: REML

- In balanced designs, $\hat{\sigma}^2_\mu$ is similar to the ANOVA-based estimation.

```
mod3 <- lmer(rating~1+(1|officerFac),data=job, REML=TRUE)
summary(mod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: rating ~ 1 + (1 | officerFac)
##    Data: job
##
## REML criterion at convergence: 145.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3841 -0.8901  0.2620  0.6496  1.2605
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  officerFac (Intercept) 80.41    8.967
##  Residual               73.28    8.561
## Number of obs: 20, groups:  officerFac, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   71.450      4.444   16.08
```

$\hat{\mu}$

① ANOVA-based (slide 21)

$\hat{\sigma}_u^2 = 80.4123$

$\hat{\sigma}^2 = 73.3$

② ML (slide 26)

$\hat{\sigma}_u^2 = 60.66$

$\hat{\sigma}^2 = 73.28$

③ REML

$\hat{\sigma}_u^2 = 80.41$

$\hat{\sigma}^2 = 73.28$

Predicting Random Effects

BLUP: Best linear unbiased predictor, find $\hat{\tau}_i$ s.t. $\min E((\hat{\tau}_i - \tau_i)^2)$

$= E((\hat{\tau}_i - E(\tau_i|Y))^2) + E((\tau_i - E(\tau_i|Y))^2)$

min when $\hat{\tau}_i = E(\tau_i|Y)$

$= E(\tau_i|\bar{Y}_{i.})$

$= \cdots$

- Specification with conditional distributions

$$Y_{ij}|\tau_i \sim_{iid} N(\mu + \tau_i, \sigma^2); \tau_i \sim_{iid} N(0, \sigma^2_\mu)$$

  - can be of interest to predict τ_i given the data
  - best predictor is $E\{\tau_i|Y\}$
- In one-way ANOVA with random effects

$$\begin{aligned} E\{\tau_i|Y\} &= E\{\tau_i|\bar{Y}_{i.}\} \\ &= E\{\tau_i\} + cov(\tau_i, \bar{Y}_{i.})[Var(\bar{Y}_{i.})]^{-1}(\bar{Y}_{i.} - E\{Y_{i.}\}) \\ &= 0 + \sigma^2_\mu \frac{1}{\sigma^2_\mu + \sigma^2/n_i}(\bar{Y}_{i.} - \mu) \end{aligned}$$

$$\boxed{\widehat{E\{\tau_i|Y\}} = \frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\mu + \hat{\sigma}^2/n_i}(\bar{Y}_{i.} - \hat{\mu})}$$

- In balanced experiments

$\bar{Y}_{i.} = \mu + \tau_i + \bar{\varepsilon}_{i.}$

$$\widehat{E\{\tau_i|Y\}} = \frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\mu + \hat{\sigma}^2/n_i}(\bar{Y}_{i.} - \hat{\mu}) \text{ Random Effects}$$

$$\hat{\tau}_i = (\bar{Y}_{i.} - \bar{Y}_{..}) \text{ ~~Random~~ Fixed Effects}$$

Extracting Parameters and Testing in R

To test

$H_0: \sigma_\mu^2 = 0 \quad H_a: \sigma_\mu^2 > 0$

① method 1: $F = \frac{MStrmt}{MSE}$

② method 2: LRT ← MLE

- Extracting values of parameters

```
fixef(mod3)    # Predicted fixed effects
vcov(mod3)     # var-cov of fixed effects
ranef(mod3)    # predicted random effects
fitted(mod3)   # Y.hat= fixed Effects hat + random effects hat
```

- Likelihood Ratio Test: $H_0 : \sigma_\mu^2 = 0 \quad vs \quad H_a : \sigma_\mu^2 > 0$
- REML-based likelihoods are not comparable for models that differ in random effects
- USE ML-based estimation
- The test is approximation since based on biased ML

```
mod4 <- lm(rating~1, data=job, REML=FALSE)
pchisq( as.numeric(2*(logLik(mod2)-logLik(mod4))),1,lower=FALSE)
```

```
## [1] 0.02919295
```

compare with P.valu: 0.0068 (F-test, Slide 21)

- Conclusion: at $\alpha = 0.05$, there is a significant variation between the officers.
- Weaker evidence than with ANOVA-based F test.

Take a break, and see you on Thursday

Example: Rating data of job applicants

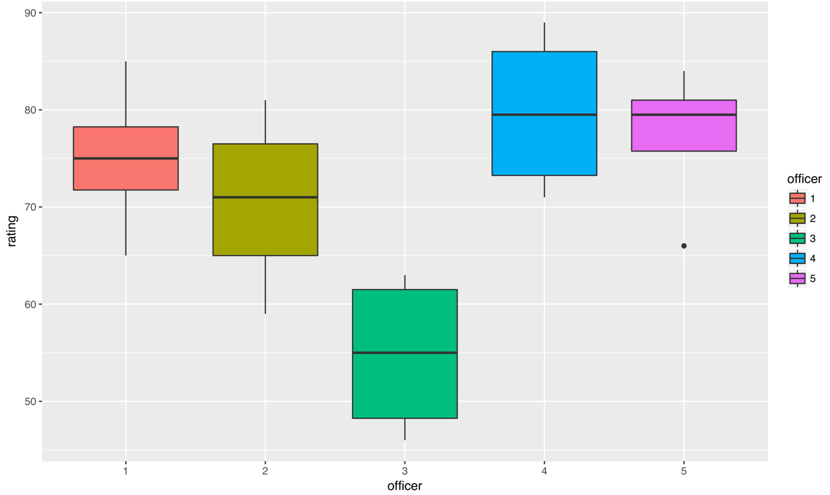
? $\mu_i = \mu + \tau_i \sim N(\mu, \sigma_\mu^2)$

Boxplot plot

① mean of rating varies with officers

② within σ^2 seems constant

```
library(ggplot2)
ggplot(job, aes(x=officer,y=rating,fill=officer)) +geom_boxplot()
```

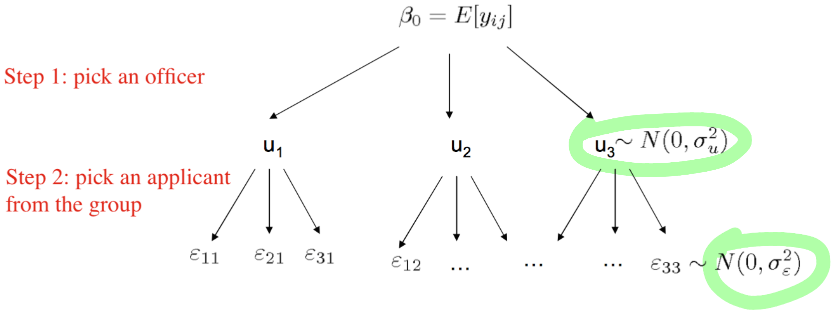


Reivew on 1-way Random Effect mdoel: Rating data of job applicants

$$Y_{ij} = \beta_0 + U_i + \epsilon_{ij},$$

where

- $U_i \sim_{iid} \ N(0, \sigma_\mu^2)$
- $\epsilon_{ij} \sim_{iid} \ N(0, \sigma^2)$
- $U \perp \epsilon$
- R command: **lmer(Rating~1+(1|officer),data=job,REML=F)**



Result: $y_{11} = \beta_0 + u_1 + \varepsilon_{11}, \ldots, y_{33} = \beta_0 + u_3 + \varepsilon_{33},$

Reivew on 1-way Random Effect mdoel: Rating data of job applicants

Mixed Effects in Matrix Notation

$$
\left[\begin{array}{c} y_{11} \\ y_{21} \\ y_{31} \\ \hline y_{12} \\ y_{22} \\ y_{32} \\ \hline y_{13} \\ y_{23} \\ y_{33} \end{array}\right]
=
\left[\begin{array}{c} 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \end{array}\right] \beta_0 +
\left[\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\right]
\left[\begin{array}{c} u_1 \\ u_2 \\ u_3 \end{array}\right] +
\left[\begin{array}{c} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \hline \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \hline \varepsilon_{13} \\ \varepsilon_{23} \\ \varepsilon_{33} \end{array}\right]
$$

- or just $Y = \overbrace{X\beta}^{\text{Fixed EF}} + \overbrace{Zu}^{\text{random EF.}} + \varepsilon$
  - β are the fixed effects
  - u are the random effects
  - X and Z are fixed, known design matrices

credit: Brian W. Junker (`brian@stat.cmu.edu`)

Example: Rating data of job applicants

mod3 = lmer(rating ~ 1+ (1| officerFac) , data =job, REML=T)

```
fixef(mod3)
```

```
## (Intercept)
##       71.45
```

$= \hat{\mu}$

```
ranef(mod3)
```

```
## $officerFac
##   (Intercept)
## 1   2.8912526
## 2  -0.7737155
## 3 -13.6011037
## 4   6.7598300
## 5   4.7237366
```

$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i$

$\hat{\tau}_i \sim N(0, \hat{\sigma}_u^2)$

Two-Way Random Effects ANOVA
(Model II in KNNL)

Car Example: KNNL 25.15 (P.1080)

- Interested in the fuel efficiency (mpg)
- Response : mpg
  - Y_{ijk} is the k-th observed mpg from driver i and car j, with $k = 1, 2, \ldots, n_{ij}$
- Two Random factors
  - Factor A: driver, levels i=1,2,. . ., 4
  - Factor B: car (same model), levels $j = 1, 2, \ldots, 5$
  - Each driver drove each car twice over the same 40-mile test course
    - $n_{ij} = 2$: balanced design.
- Scientific question:
  - How much of the $\underbrace{\text{overall variability}}_{\sigma_Y^2}$ is due to driver and/or car?

σ_D^2 σ_C^2

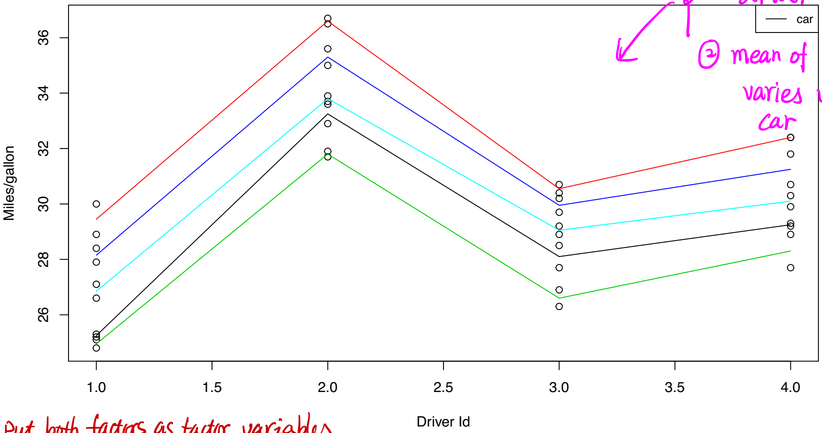
Car example

Mean plot

```
dat2515 <- read.table("CH25PR15.txt",sep="",as.is=T,header=F)
colnames(dat2515) = c("mpg","driver","car","replicates")
plot(dat2515$driver, dat2515$mpg,xlab="Driver Id",ylab="Miles/gallon")
m <- with(dat2515, tapply(mpg, list(car,driver),mean))
for (i in 1:5) { lines(1:4, m[i,],col=i )}
legend("topright",lty=1,"car",cex=0.8)
```

① Mean of Y varies with driver

② mean of Y varies with car



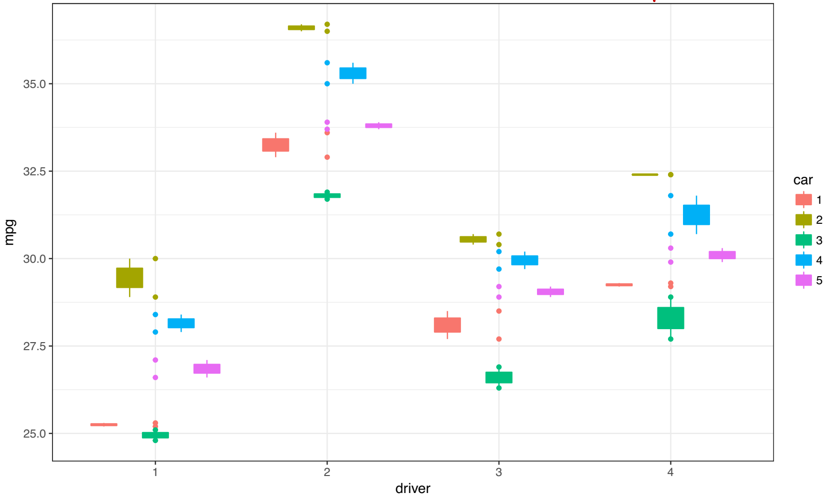
Put both factors as factor variables

```
dat2515$car=as.factor(dat2515$car); dat2515$driver=as.factor(dat2515$driver)
```

Car example: Boxplot

```
ggplot(dat2515, aes(driver, mpg, fill=car, color=car)) +
    geom_boxplot() + theme_bw(base_size=12)+geom_point()
```

give similar infor.
as from mean plot



Random factor effects model

- In mathematical notation

$$\boxed{Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}}$$

$$i = 1, 2, \ldots, a; \quad j = 1, \ldots b, \quad k = 1, \ldots n$$

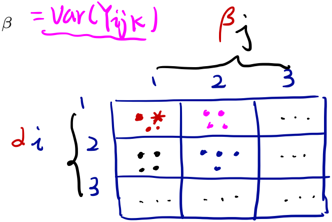
same model form as two-way ANOVA with interaction term.

- μ : grand mean (or another reference)
- α_i: random deviation of the i-th level of factor A from the reference. $\alpha_i \sim_{iid} \; N(0, \sigma^2_\alpha)$
- β_j: random deviation of the j-th level of factor B from the reference. $\beta_j \sim_{iid} \; N(0, \sigma^2_\beta)$
- $(\alpha\beta)_{ij}$: the joint random effect of the i-th level of factor A and j-th level of b. $(\alpha\beta)_{ij} \sim_{iid} \; N(0, \sigma^2_{\alpha\beta})$
- ϵ_{ijk} : random error, $\epsilon_{ijk} \sim_{iid} \; N(0, \sigma^2)$
- All random terms are independent.

Covariance Structure

$$Y_{ijk} = \mu + \overset{\sigma_\alpha^2}{\alpha_i} + \overset{\sigma_\beta^2}{\beta_j} + \overset{\sigma_{\alpha\beta}^2}{(\alpha\beta)_{ij}} + \overset{\sigma^2}{\epsilon_{ijk}} \rightarrow Y_{ijk} \sim N(\mu, \text{sum } \sigma_i^2)$$

- There are four variance/covariance parameters
  - $cov(Y_{ijk}, Y_{ijk}) = \sigma^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2$ $= Var(Y_{ijk})$
  - $cov(Y_{ijk}, Y_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2$
  - $cov(Y_{ijk}, Y_{ij'k'}) = \sigma_\alpha^2$
  - $cov(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2$
  - $cov(Y_{ijk}, Y_{i'j'k}) = 0$



- Y_{ijk} are not independent
- Questions
  - percentage of total variation due to each factor
  - percentage of cell means variation (i.e. ignoring error variance)
  - pairwise comparisons between levels of factors are not appropriate.

ANOVA Table

*

Source	d.f.	EMS
A	a-1	$\sigma^2 + n\sigma^2_{\alpha\beta} + bn\sigma^2_\alpha$
B	b-1	$\sigma^2 + n\sigma^2_{\alpha\beta} + an\sigma^2_\beta$
AB	(a-1)(b-1)	$\sigma^2 + n\sigma^2_{\alpha\beta}$
Error	ab(n-1)	σ^2
Total	abn-1	

difference

$bn\sigma_\alpha^2$

$an\sigma_\beta^2$

$n\sigma_{\alpha\beta}^2$

- Parameter estimates using mean squares
  - $\hat{\sigma}^2 = MSE$
  - $\hat{\sigma}^2_{\alpha\beta} = (MSAB - MSE)/n$
  - $\hat{\sigma}^2_\beta = (MSB - MSAB)/an$
  - $\hat{\sigma}^2_\alpha = (MSA - MSAB)/n$
- Estimates can be negative.
- Same procedure for CI and same adjustments for df, as in the one-way random effects ANOVA.

ANOVA based hypothesis tests

Three tests of variance, no hierarchy

- $H_{0AB} : \sigma^2_{AB} = 0$ vs $H_{1AB} : \sigma^2_{\alpha\beta} > 0$

$$F = \frac{MSAB}{MSE} \sim_{H_0} F_{(a-1)(b-1),ab(n-1)}$$

✓ consistent with F test in ANOVA.

- $H_{0A:}\sigma^2_{\alpha} = 0$ vs $H_{1AB} : \sigma^2_{\alpha} > 0$

$$F = \frac{MSA}{MSAB} \sim_{H_0} F_{(a-1),(a-1)(b-1)}$$

- $H_{0B:}\sigma^2_{\beta} = 0$ vs $H_{1B} : \sigma^2_{\beta} > 0$

$$F = \frac{MSB}{MSAB} \sim_{H_0} F_{(b-1),(a-1)(b-1)}$$

Hower, in ANOVA Table,

$F = \frac{\text{---}}{MSE}$

Car example: ANOVA-based Approach

```
carm1 <- aov(mpg~driver*car,data=dat2515)
summary(carm1)
```

MSA → MSB

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## driver       3 280.28   93.43  531.60  < 2e-16 ***
## car          4  94.71   23.68  134.73 3.66e-14 ***
## driver:car  12   2.45    0.20    1.16    0.371
## Residuals   20   3.52    0.18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MSAB

→ $H_0: \sigma^2_{AB} = 0$ $H_a: \sigma^2_{AB} > 0$

```
# Estimate variance components
( sigma2ABhat <- (0.204-0.176)/2 )
```

```
## [1] 0.014
```

← $\hat{\sigma}^2_{AB} = \frac{MSAB - MSE}{n}$

```
( sigma2DriverHat <- (93.428-0.204)/(2*5) )
```

```
## [1] 9.3224
```

← $\hat{\sigma}^2_{A} = \frac{MSA - MSAB}{bn}$

```
( sigma2ABhat <- (23.678-0.204)/(2*4) )
```

```
## [1] 2.93425
```

← $\hat{\sigma}^2_{B} = \frac{MSB - MSAB}{an}$

- Only the F test for the interaction is used directly. ✓

Car example: REML-based Approach

```r
carm2 <- lmer(mpg~1+(1|driver)+(1|car)+(1|driver:car),data=dat2515, REML=T)
summary(carm2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mpg ~ 1 + (1 | driver) + (1 | car) + (1 | driver:car)
##    Data: dat2515
##
## REML criterion at convergence: 86.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.54828 -0.61813 -0.08972  0.61864  1.98982
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  driver:car (Intercept) 0.01406  0.1186
##  car        (Intercept) 2.93431  1.7130
##  driver     (Intercept) 9.32243  3.0533
##  Residual               0.17575  0.4192
## Number of obs: 40, groups:  driver:car, 20; car, 5; driver, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    30.05       1.71   17.58
```

$\hat{\sigma}^2_{AB}$ (REML)

$\hat{\sigma}^2_B$

$\hat{\sigma}^2_A$

$\hat{\sigma}^2$

$\hat{\mu} =$

Car example: ML-based Approach

```r
carm3 <- lmer(mpg~1+(1|driver)+(1|car)+(1|driver:car),data=dat2515, REML=FALSE)
summary(carm3)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mpg ~ 1 + (1 | driver) + (1 | car) + (1 | driver:car)
##    Data: dat2515
##
##      AIC      BIC   logLik deviance df.resid
##     99.6    108.0    -44.8     89.6       35
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.54685 -0.61866 -0.09137  0.61441  1.98837
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  driver:car (Intercept) 0.01408  0.1187
##  car        (Intercept) 2.77213  1.6650
##  driver     (Intercept) 7.41376  2.7228
##  Residual               0.17575  0.4192
## Number of obs: 40, groups:  driver:car, 20; car, 5; driver, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   30.048      1.553   19.34
```

		$\hat{\sigma}^2_{AB}$	D σ^2_A	Car σ^2_B
P.45	ANOVA	0.014	9.3224	2.9342
P.46	REML	0.01406	9.32243	2.93431
P.47.	MLE	0.01408	7.41376	2.77213

Car example: ML-based Approach

```
confint(carm2) # REML based
```

```
##                  2.5 %     97.5 %
## .sig01       0.0000000  0.4164471
## .sig02       0.9554686  3.8124720
## .sig03       1.5358586  6.7100201
## .sigma       0.3165499  0.5618101
## (Intercept) 26.3641083 33.7308929
```

CI contains 0

$\to \sigma^2_{AB} = 0$

same. using "profile method"

```
confint(carm3) # ML based
```

```
##                  2.5 %     97.5 %
## .sig01       0.0000000  0.4164471
## .sig02       0.9554686  3.8124720
## .sig03       1.5358586  6.7100201
## .sigma       0.3165499  0.5618101
## (Intercept) 26.3641083 33.7308929
```

$\to \sigma^2_{AB} = 0$

Car example: REML additive model

```
carm4 <- lmer(mpg~1+(1|driver)+(1|car),data=dat2515, REML=TRUE)
summary(carm4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mpg ~ 1 + (1 | driver) + (1 | car)
##    Data: dat2515
##
## REML criterion at convergence: 86.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5426 -0.6016 -0.1037  0.6160  2.0356
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  car      (Intercept) 2.9365   1.7136
##  driver   (Intercept) 9.3242   3.0536
##  Residual             0.1863   0.4316
## Number of obs: 40, groups:  car, 5; driver, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    30.05       1.71   17.57
```

- SAS implementation produces REML estimates that are identical to ANOVA but R gives slightly different result.
- This is due to differences in the implementation of REML between SAS and R.

Car example: ML additive model

```r
carm5 <- lmer(mpg~1+(1|driver)+(1|car),data=dat2515, REML=FALSE)
summary(carm5)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mpg ~ 1 + (1 | driver) + (1 | car)
##    Data: dat2515
##
##      AIC      BIC   logLik deviance df.resid
##     97.7    104.4    -44.8     89.7       36
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5411 -0.6018 -0.1011  0.6117  2.0342
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  car      (Intercept) 2.7744   1.6656
##  driver   (Intercept) 7.4158   2.7232
##  Residual             0.1863   0.4316
## Number of obs: 40, groups:  car, 5; driver, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   30.048      1.554   19.34
```

Car example: CI based REML/ML additive model

```
# default confint(lmer.obj, method="profile")
confint(carm4) # REML based
```

```
##                   2.5 %     97.5 %
## .sig01        0.9567397  3.8128039
## .sig02        1.5365191  6.7103803
## .sigma        0.3441403  0.5636012
## (Intercept) 26.3637911 33.7312118
```

same

```
confint(carm5) # ML based
```

```
##                   2.5 %     97.5 %
## .sig01        0.9567397  3.8128039
## .sig02        1.5365191  6.7103803
## .sigma        0.3441403  0.5636012
## (Intercept) 26.3637911 33.7312118
```

different

```
confint(carm4,method="boot")
```

→ bootstrap method: simulation method.

```
##                   2.5 %     97.5 %
## .sig01        0.6401579  2.8921306
## .sig02        0.7165444  5.4011820
## .sigma        0.3158970  0.5283283
## (Intercept) 26.3655010 33.2013851
```

```
#confint(m1,method="boot",boot.type="norm")
#confint(m1,method="boot",boot.type="basic")
#confint(m1,method="boot",boot.type="perc")
```

After Lecture This Week

Practice problems

- Review all the slides
- Try all the R example in slides.
- Try all posted practice problem (solution is available)

Topics for next week:

- Topic in the following week after midterm
  - Two-way mixed effects ANOVA (Model III in KNNL)