Fixed vs. Random Effects

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

Outline

- Review One-Way ANOVA
- Fixed vs. Random Effects
- Random Effects

ANOVA Review (KNN Chap 16)

Analysis of variance (ANOVA) models are regression models with qualitative predictors, called factors or treatments.

Factors have different levels.

For example, the factor "education" may have the levels *high school, undergraduate, graduate*. The factor "gender" has two levels *female, male*.

We may have several factors as predictors, e.g. race and gender may be used to predict annual salary in $.

16.3 Cell means model

Have r different treatments or factor levels. At each level i, have n_i observations from group i.

Total number of observations is $n_T = n_1 + n_2 + \cdots + n_r$.

Response is Y_{ij} where $\left\{ \begin{array}{ll} i = 1, \dots, r & \text{factor level} \\ j = 1, \dots, n_i & \text{obs. within factor level} \end{array} \right\}$.

Example: Two factors: MS, PhD. Y_{ij} is age in years. Spring of 2014 we observe

$$Y_{11} = 28,\ Y_{12} = 24,\ Y_{13} = 24,\ Y_{14} = 22,\ Y_{15} = 26,\ Y_{16} = 23,$$

$$Y_{21} = 29,\ Y_{22} = 23,\ Y_{23} = 26,\ Y_{24} = 25,\ Y_{25} = 22,\ Y_{26} = 23,\ Y_{27} = 38,\ Y_{28} = 33,\ Y_{29} = 30,\ Y_{2,10} = 27.$$

One-way ANOVA model

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

Can rewrite as

$$Y_{ij} \overset{ind.}{\sim} N(\mu_i, \sigma^2).$$

- Data are normal, data are independent, variance constant across groups.
- μ_i is allowed to be different for each group. $\mu_1, \ldots, \mu_r$ are the r population means of the response. A picture helps.
- Questions: what is $E\{Y_{ij}\}$? What is $\sigma^2\{Y_{ij}\}$?

Matrix formulation

(pp. 683–684, 710–712 in KNN) For $r = 3$ we have

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n_2} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n_2} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3n_3} \end{bmatrix}$$

or

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon.$$

For $r = 3$, let $Q(\mu_1, \mu_2, \mu_3) = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2$.

Need to minumize this over all possible (μ_1, μ_2, μ_3) to find least-squares (LS) solution. Can easily show that $Q(\mu_1, \mu_2, \mu_3)$ has minimum at

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1\cdot} \\ \bar{Y}_{2\cdot} \\ \bar{Y}_{3\cdot} \end{bmatrix}$$

where $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ is the sample mean from the ith group (pp. 687–688).

These $\hat{\boldsymbol{\beta}}$ are also maximum likelihood estimates.

Matrix formula of least-squares estimators ($r = 3$)

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} n_1^{-1} & 0 & 0 \\ 0 & n_2^{-1} & 0 \\ 0 & 0 & n_3^{-1} \end{bmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} Y_{1\cdot} \\ Y_{2\cdot} \\ Y_{3\cdot} \end{bmatrix},$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \bar{Y}_{1\cdot} \\ \bar{Y}_{2\cdot} \\ \bar{Y}_{3\cdot} \end{bmatrix}.$$

Residuals

As in regression (STAT 704),

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i = Y_{ij} - \bar{Y}_{i\cdot}$$

As usual, $\hat{Y}_{ij}$ is the estimated mean response under the model.

Note that $\sum_{j=1}^{n_i} e_{ij} = 0, \; i = 1, \ldots, r$. [check this!]

In matrix terms

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\beta} = \mathbf{Y} - \hat{\mathbf{Y}}.$$

Define the following

$$Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij} = i \text{ group sum},$$

$$\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = i\text{th group mean}$$

$$Y_{\cdot\cdot} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{r} Y_{i\cdot} = \text{sum all obs.}$$

$$\bar{Y}_{\cdot\cdot} = \frac{1}{n_T} \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_T} \sum_{i=1}^{r} Y_{i\cdot} = \text{mean all obs.}$$

Sums of squares for treatments, error, and total

$$
\begin{aligned}
\text{SSTO} &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = \text{variability in } Y_{ij}\text{'s} \\
\text{SSTR} &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\hat{Y}_{ij} - \bar{Y}_{\cdot\cdot})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\hat{\mu}_{ij} - \bar{Y}_{\cdot\cdot})^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum_{i=1}^{r} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \\
&= \text{variability explained by ANOVA model} \\
\text{SSE} &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \hat{Y}_{ij})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} e_i^2 \\
&= \text{variability NOT explained by ANOVA model}
\end{aligned}
$$

Comments

- As before in regression,

$$\underbrace{\text{SSTO}}_{\text{total}} = \underbrace{\text{SSTR}}_{\text{treatment effects}} + \underbrace{\text{SSE}}_{\text{leftover randomness}}$$

- SSE=0 $\Rightarrow Y_{ij} = Y_{ik}$ for all $j \neq k$ and for $i = 1, \dots, r$.
- SSTR=0 $\Rightarrow \bar{Y}_{i\cdot} = \bar{Y}_{\cdot\cdot}$ for $i = 1, \dots, r$.

ANOVA table (p. 694)

Source	SS	df	MS	E(MS)
SSTR	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	$\text{SSTR}/(r-1)$	$\sigma^2 + \frac{\sum_{i=1}^{r} n_i(\mu_i - \mu_\cdot)^2}{r-1}$
SSE	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$n_T - r$	$\text{SSE}/(n_T - r)$	σ^2
SSTO	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n_T - 1$		

Degrees of freedom

- SSTO has $n_T - 1$ df because there are n_T $Y_{ij} - \bar{Y}_{..}$ terms in the sum, but they add up to zero (1 constraint).
- SSE has $n_T - r$ df because there are n_T $Y_{ij} - \bar{Y}_{i.}$ terms in the sum, but there are r constraints of the form $\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.}) = 0,\ i = 1, \ldots, r$.
- SSTR has $r - 1$ df because there are r terms $n_i(\bar{Y}_{i.} - \bar{Y}_{..})$ in the sum, but they sum to zero (1 constraint).

Assuming $\mu_1 = \cdots = \mu_r$, Cochran's Theorem (Section 2.7) shows that $SSTR/\sigma^2 \sim \chi^2_{r-1}$ and $SSE/\sigma^2 \sim \chi^2_{n_T - r}$ and they are independent.

Expected mean squares

$$E\{\text{MSE}\} = \sigma^2, \quad \text{MSE is unbiased estimate of } \sigma^2$$

$$E\{\text{MSTR}\} = \sigma^2 + \frac{\sum_{i=1}^{r} n_i(\mu_i - \mu_.)^2}{r-1},$$

where $\mu_. = \sum_{i=1}^{r} \frac{n_i \mu_i}{n_T}$ is the weighted average of $\mu_1, \ldots, \mu_r$ (pp. 696–698).

If $\mu_i = \mu_j$ for all $i, j \in \{1, \ldots, r\}$ then $E\{\text{MSTR}\} = \sigma^2$, otherwise $E\{\text{MSTR}\} > \sigma^2$.

Hence, if any group means are different then $\frac{E\{\text{MSTR}\}}{E\{\text{MSE}\}} > 1$.

16.6 F test of $H_0 : \mu_1 = \cdots = \mu_r$

Fact: If $\mu_1 = \cdots = \mu_r$ then

$$F^* = \frac{\text{MSTR}}{\text{MSE}} \sim F(r-1, n_T - r).$$

To perform α-level test of $H_0 : \mu_1 = \cdots = \mu_r$ vs. H_a : some $\mu_i \neq \mu_j$ for $i \neq j$,

- Accept if $F^* \leq F(1-\alpha, r-1, n_T - r)$ or p-value $\geq \alpha$.
- Reject if $F^* > F(1-\alpha, r-1, n_T - r)$ or p-value $< \alpha$.

p-value $= P\{F(r-1, n_T - 1) \geq F^*\}$.

Comments

- If $r = 2$ then $F^* = (t^*)^2$ where t^* is t-statistic from 2-sample pooled-variance t-test.
- The F-test may be obtained from the general nested linear hypotheses approach (big model / little model). Here the full model is $Y_{ij} = \mu_i + \epsilon_{ij}$ and the reduced is $Y_{ij} = \mu + \epsilon_{ij}$.

$$F^* = \frac{\left[\frac{SSE(R) - SSE(F)}{dfE_R - dfE_F}\right]}{\frac{SSE(F)}{dfE_F}} = \frac{MSTR}{MSE}.$$

An alternative formula for ANOVA model can be written as:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where $\alpha_r = 0$.

- $E\{Y_{rj}\} = \mu$; μ is the cell-mean for the rth level.
- For $i < r$, $E\{Y_{ij}\} = \mu + \alpha_i$; α_i is i's offset to group r's mean μ.

Can we express this formulation using a linear regression model?

Fixed vs. Random Effects

- In ANOVA, the categorical variable are well-defined categories: such as genotype groups, age groups
- In some designs, the categorical variable is "subject."
- Simplest example: repeated measures, where more than one measurement is taken on the same individual.
- In this case, the "group" effect α_i is best though of as random because we only sample a subset of the entire population of subjects.

When to use random effects

- A “group” effect is random if we can think of the levels we observe in the group to be samples from a larger population.
- Example: if collecting data from different medical centers, “center” might be thought of as random.
- Example: if surveying students on different campuses, “campus” may be a random effect.

Examples: Sodium content in beer

- How much sodium is there in North American beer? How much does this vary by brand?
- Observations: for 6 brands of beer, researchers recorded the sodium content of 8 12-ounce bottles
- Questions of interest: what is the "grand mean" sodium content? How much variability is there from brand to brand?
- Since brad was sampled from a larger set, we can think of various brands as random variables.
- "Individuals" in this case are brands, repeated measured 8 times.

25.1 One-way random cell means model

If treatment levels come from a larger population, their effects are best modeled as random. A one-way random cell means model is

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where

$$\mu_1, \ldots, \mu_r \overset{iid}{\sim} N(\mu_\cdot, \sigma_\mu^2) \text{ independent of } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

As usual, $i = 1, \ldots, r$ and $j = 1, \ldots, n_i$.

The test of interest is $H_0 : \sigma_\mu^2 = 0$.

We can re-express the model as a random effects model, by writing $\mu_i = \mu_\cdot + \tau_i$, where $\tau_1, \ldots, \tau_r \overset{iid}{\sim} N(0, \sigma_\mu^2)$.

$\tau_1, \ldots, \tau_r$ are called *random effects* and σ_μ^2 and σ^2 are termed *variance components*. This model is an example of a *random effects* model, because it has only random effects beyond the intercept $\mu_\cdot$ (which is fixed).

Model properties

The random cell means model has some quite different properties from the fixed cell means model.

1. $E\left(Y_{ij}\right) = \mu.$
2. $\sigma^2\left\{Y_{ij}\right\} = \sigma^2 + \sigma_\mu^2$ (Hence the term *variance components*)
3. $\sigma\left\{Y_{ij}, Y_{ij'}\right\} = \sigma_\mu^2$
4. $\rho\left\{Y_{ij}, Y_{ij'}\right\} = \frac{\sigma_\mu^2}{\sigma^2+\sigma_\mu^2}$
5. $E\left(\bar{Y}_{..}\right) = \mu.$
6. $\sigma^2\left(\bar{Y}_{..}\right) = \frac{\sigma^2+n\sigma_\mu^2}{rn}$

Variance-Covariance Matrix

- Suppose r=2 levels, and n=2 cases in each level. The observation vector is:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix}.$$

- The variance-covariance matrix of $\mathbf{Y}$ is:

Variance-Covariance Matrix

- Suppose r=2 levels, and n=2 cases in each level. The observation vector is:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix}.$$

- The variance-covariance matrix of $\mathbf{Y}$ is:

$$Cov(\mathbf{Y}) = \begin{bmatrix} \sigma^2 + \sigma_\mu^2 & \sigma_\mu^2 & 0 & 0 \\ \sigma_\mu^2 & \sigma^2 + \sigma_\mu^2 & 0 & 0 \\ 0 & 0 & \sigma^2 + \sigma_\mu^2 & \sigma_\mu^2 \\ 0 & 0 & \sigma_\mu^2 & \sigma_\mu^2 + \sigma^2 \end{bmatrix}.$$

Simple random effect model

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where

$$\mu_1, \ldots, \mu_r \stackrel{iid}{\sim} N(\mu_{\cdot}, \sigma^2_\mu)$$

independent of

$$\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2).$$

As usual, $i = 1, \ldots, r$ and $j = 1, \ldots, n_i$.

- We might be interested in the population mean μ: CI, is it zero?
- What is really usually the focus: σ_μ: CI, is it zero?

ANOVA table for one-way random effect

Source	SS	df	MS	E(MS)
SSTR	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	SSTR/$(r-1)$	$\sigma^2 + n\sigma_\mu^2$
SSE	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$n_T - r$	SSE/$(n_T - r)$	σ^2
SSTO	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n_T - 1$		

- Only change here is the expectation of MSTR reflects the randomness of μ_is
- Under $H_0 : \sigma_\mu^2 = 0$, it is easy to see that

$$\frac{MSTR}{MSE} \sim F_{(r-1),(n-1)r}$$

Inference for μ

- We know that $E(\overline{Y_{..}}) = \mu$ and $\sigma^2(\overline{Y_{..}}) = \frac{n\sigma_\mu^2 + \sigma^2}{rn}$.
- Therefore,

$$\frac{\overline{Y_{..}} - \mu_.}{\sqrt{\frac{SSTR}{(r-1)rn}}} \sim t_{r-1}$$

- Why $r - 1$ degree of freedom? If we can sample infinite number of observations fro each level, so that $\overline{Y_{i.}} \to \mu_i$
- To learn anything about μ, we still only have r observations $(\mu_1, \mu_2, ...\mu_r)$
- Hence, sampling more within the "group" can not narrow the CI for $\mu_.$

Testing $H_0 : \sigma_\mu = 0$

The MSE and MSTR are defined as they were before. One can show $E(MSE) = \sigma^2$ and $E(MSTR) = \sigma^2 + n\sigma_\mu^2$ when $n = n_i$ for all i. Most packages provides symbolic forms of expected mean squares for random/mixed models if requested.

If $\sigma_\mu = 0$ we expect $F^* = MSTR/MSE$ to be somewhat larger than 1. In fact, just like the fixed-effects case, $F^* \sim F(r-1, n_T - r)$. This is the test given by `proc glm` when you add a `random A;` statement.

One can also fit the model in `proc mixed`, but this procedure provides a slightly cruder test of $H_0 : \sigma_\mu = 0$.

Disadvantages of ANOVA estimators

$$\widehat{\sigma^2} = SSE/r[n-1] = MSE, \quad \widehat{\sigma^2_\mu} = (MSTR - MSE)/n$$

- When MSTR < MSE, $\sigma^2_\mu < 0$, this is rather embarrassing.
- The solution is not unique in unbalanced designs.
- The need for complicated algebraic calculations in more complex designs.

Other tests and estimates

We can derive estimates for $\mu.$, σ^2 and $\frac{\sigma_\mu^2}{\sigma^2+\sigma_\mu^2}$ because pivotal quantities are readily available. It is an open question whether we are interested in inference on $\mu.$ in most practical applications.

Other quantities of interest tended to require moment-based estimates (old school)–e.g., the variance component σ_μ^2. Methods to provide point estimates and/or standard errors include

- Maximum Likelihood (biased)
- Restricted Maximum Likelihood

ML estimate

The log-likelihood for a simple linear regression model is:

$$
\begin{aligned}
l(\beta, \sigma, \sigma_\mu | Y, X) &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\sigma^2| - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta) \\
\widehat{\sigma^2} &= \frac{(Y - X\beta)^T(Y - X\beta)}{n} \\
E(\sigma^2) &= \frac{n-1}{n}\sigma^2.
\end{aligned}
$$

ML estimate is biased because of the unknown estimator for the mean!

REML

The log-likelihood for the data is:

$$l(\beta, \sigma, \sigma_\mu | Y, X) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)$$

Integrate the log-likelihood w.r.t β in REML:

$$l(\beta, \sigma, \sigma_\mu | Y, X) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| + \log\left[\int e^{-\frac{(Y-X\beta)^T\Sigma^{-1}(Y-X\beta)}{2}} d\beta\right]$$

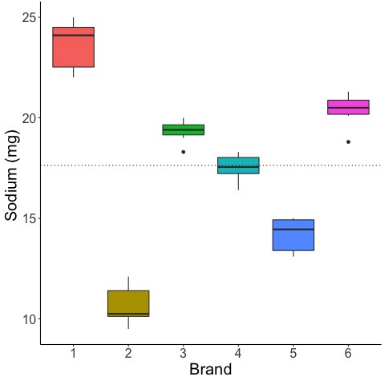
Let $f(\beta) = -\frac{(Y-X\beta)^T\Sigma^{-1}(Y-X\beta)}{2}$; use Taylor expansion:

$$\begin{aligned}
f(\beta) &\approx f(\widehat{\beta}) + \frac{1}{2}(\beta - \widehat{\beta})^2 f''(\widehat{\beta}) \quad \text{Note that}[f'(\beta) = 0] \\
f(\beta) = -\frac{(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)}{2} &\approx -\frac{(Y - X\widehat{\beta})^T\Sigma^{-1}(Y - X\widehat{\beta})}{2} - \frac{(\beta - \widehat{\beta})^T X^T\Sigma^{-1}X(\beta - \widehat{\beta})}{2} \\
\log\left[\int L(\sigma, \sigma_\mu | Y, X)\right] &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{(Y - X\widehat{\beta})^T\Sigma^{-1}(Y - X\widehat{\beta})}{2} \\
&+ \log\left[\int e^{-\frac{(\beta-\widehat{\beta})^T X^T\Sigma^{-1}X(\beta-\widehat{\beta})}{2}} d\beta\right] \text{ Laplace appromixation} \\
\log\left[\int L(\sigma, \sigma_\mu | Y, X)\right] &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(Y - X\widehat{\beta})^T\Sigma^{-1}(Y - X\widehat{\beta}) - \frac{1}{2}\log(|X^T\Sigma^{-1}X|)
\end{aligned}$$

REML does not depend on β!

Sodium content in beer

	sodium	brand	rep
1	24.4	1	1
2	22.6	1	2
3	23.8	1	3
4	22.0	1	4
⋮			
45	20.1	6	5
46	18.8	6	6
47	21.1	6	7
48	20.3	6	8



One-Way ANOVA, fixed effect

```
 > summary(aov(sodium ~ brand))
            Df Sum Sq Mean Sq F value Pr(>F)
brand        5  854.5  170.91   238.7 <2e-16 ***
Residuals   42   30.1    0.72
```

$\widehat{\mu} = 17.62, \quad \widehat{\sigma^2} = 0.72, \; \widehat{\sigma_\mu} = (170.91 - 0.72)/8 = 21.27$

Random effect model (REML)

The original R pacakge was `nlme` as described in Pinheiro and Bates (2000). Subsequently Bates (2005) introduced the package `lme4`.

```
 > library(lme4)
Loading required package: Matrix
> mmod<-lmer(sodium ~ 1 + (1|brand), data=beer)
> summary(mmod)
Linear mixed model fit by REML ['lmerMod']
Formula: sodium ~ 1 + (1 | brand)
   Data: beer

REML criterion at convergence: 148.9

Random effects:
 Groups   Name        Variance Std.Dev.
 brand    (Intercept) 21.274   4.6123
 Residual              0.716   0.8461
Number of obs: 48, groups:  brand, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)   17.629      1.887   9.343
```

ML estimates

```
> smod<-lmer(sodium ~ 1 + (1|brand), data=beer, REML=F)
> summary(smod)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: sodium ~ 1 + (1 | brand)
   Data: beer

     AIC      BIC   logLik deviance df.resid
   157.9    163.6    -76.0    151.9       45

Random effects:
 Groups   Name        Variance Std.Dev.
 brand    (Intercept) 17.713   4.2087
 Residual              0.716   0.8461
Number of obs: 48, groups:  brand, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)   17.629      1.723   10.23
```

$\sigma_\mu^2 = 17.713$, ML estimate biases toward zero! Fixed effects remain the same.

Likelihood ratio test

$$2[l(\widehat{\beta_1}, \widehat{\sigma_1}, \widehat{\sigma_{\mu_1}}|y, x) - l(\widehat{\beta_0}, \widehat{\sigma_0}, \widehat{\sigma_{\mu_0}}|y, x)]$$

- For testing fixed effects, we cannot use the REML estimation approach. Use ordinary ML instead.
- This test statistic is approximately chi-squared with degrees of freedom equal to the difference in the dimensions of the two parameters spaces
- Unfortunately, this test requires several assumptions (parameters under the null are not on the boundary). Serious problems can arise with this approximation.
- The *p values* for the fixed effects ten to be too small and the *p values* for the random effects tend to be too large.

F-test

- In F-test for fixed effect, the definition of degree of freedom becomes murky in the presence of random effect parameters.
- For simple models with balanced data, the F-test is correct but in more complex models or unbalanced data, *p values* can be substantially incorrect. For this reason, `lme4` declines to state *p values*.
- The t-statistics also rely on the same problematic approximations.

Model selection

$$AIC = -2\max\left(\log \text{likelihood}\right) + 2p$$

- Okay to use when compare fixed effect parameters as the number of random effect will be the same
- Comparing models with varying random effects is problematic due to the boundary issue.

Likelihood ratio test

```
> nullmod<-lm(sodium ~ 1, data=beer)
> anova(smod, nullmod)
Data: beer
Models:
nullmod: sodium ~ 1
smod: sodium ~ 1 + (1 | brand)
        npar    AIC    BIC   logLik deviance  Chisq Df
nullmod    2 280.09 283.83 -138.043   276.09
smod       3 157.94 163.55  -75.968   151.94 124.15  1
        Pr(>Chisq)
nullmod
smod     < 2.2e-16 ***
```

Parametric bootstrap

```
> library("faraway")
> sim<-1000
> lrtstat<-rep(NA, sim)
> for(i in 1:sim){
+  y<-unlist(simulate(nullmod))
+  bnull<-lm(y ~ 1)
+  balt<-lmer(y ~ 1 + (1|brand), data=beer, REML=F)
+  lrtstat[i]<-as.numeric(2*(logLik(balt)-logLik(bnull)))
+ }
> mean(lrtstat> obslrt)
[1] 0
```

Parametric boostrap

```
> library("RLRsim")
> nullmod<-lm(sodium ~ 1, data=beer)
> exactLRT(smod, nullmod)
No restrictions on fixed effects. REML-based inference preferable.

simulated finite sample distribution of LRT. (p-value
based on 10000 simulated values)

data:
LRT = 124.15, p-value < 2.2e-16
> exactRLRT(mmod)

simulated finite sample distribution of RLRT.

(p-value based on 10000 simulated values)

data:
RLRT = 126.27, p-value < 2.2e-16
```

Predict random effect (α_i)

```
> ranef(mmod)
$brand
  (Intercept)
1   5.9831634
2  -6.9250345
3   1.7011768
4  -0.1286256
5  -3.4023537
6   2.7716735
with conditional variances for \brand"
> round(tapply(sodium,brand,mean)-mean(sodium),3)
     1      2      3      4      5      6
 6.008 -6.954  1.708 -0.129 -3.417  2.783
```

- The model parameters in the random effect model are $\mu, \sigma^2, \sigma_\mu^2, \alpha_i$ is considered as model parameters but just a random realization from the population of α_i.

Shrinkage estimates

```
> fit2<-lm(sodium ~ brand-1)
> coef(fit2)
 brand1  brand2  brand3  brand4  brand5  brand6
23.6375 10.6750 19.3375 17.5000 14.2125 20.4125
> r<-coef(fit2)-mean(coef(fit2))
> r
    brand1     brand2     brand3     brand4     brand5
 6.0083333 -6.9541667  1.7083333 -0.1291667 -3.4166667
    brand6
 2.7833333
> rr<-ranef(mmod)
> rr$brand/r
  (Intercept)
1   0.9958108
2   0.9958108
3   0.9958108
4   0.9958108
5   0.9958108
6   0.9958108
```

Fixed vs. random effects

- Fixed effects are constant values but random effects follow a distribution.
- Effects are fixed if they are the interest, ie the fixed α_i or random if there is interest in the underlying population with variance estimate σ^2_μ
- When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random.
- Fixed effects are estimated using least squares and random effects are estimated with shrinkage like REML.