STAT 3119

Week 13: 11/21/2019 @GWU

Outline

- Two factor ANOVA model II-III
- Model and estimation with examples (ch 25.2-25.4)

Recall : Fixed vs. random effects models

- **Fixed** effects models, in which the levels of each factor were fixed in advance of the experiment and we were interested in differences in response among those **specific** levels used in the given study.
- **Random** effects models, in which the factor levels are a sample from a larger population of potential factor levels. We are interested in whether that the factors have significant effects in explaining the response for the general population with many possible levels for the factors.

Two-Factor Studies- ANOVA model II and III

Data for two-way design

- Y, the response variable,
- Factor A with levels $i = 1, ..., a$,
- Factor B with levels $j = 1, ..., b$,
- Y_{ijk} is the kth observation in factor level (i, j), $k = 1, ...n_{ij}$
- For balanced designs, $n_{ij} = n$

Example 1: ANOVA Model II- random effects model

Consider an investigation of the effects of machine operators (factor A) and machines (factor B) on the number of pieces produced in a day.

- Five operators and three machines are used in the study. Yet the inferences are not to be confined to the particular five operators and three machines participating in the study, but rather they are to pertain to all operators and all machines available to the company.
- Here a random factor effects ANOVA model (model II) would be appropriate for the two-factor study if each of the two sets of factor levels may be considered the result of sampling a population (all operators, all machines) about which inferences are to be drawn.

Example 2: ANOVA Model III- Mixed effects model

A investigation of the effects of four different training methods (factor A) and five instructors (factor B) upon learning in a company training program.

- The four levels for training methods may be considered fixed, since interest centers in these particular training methods.
- The levels for instructors may be viewed as random, since inferences are to be made about a population of instructors of which the five used in the study are viewed as a sample.

Two-Factor Studies: ANOVA Model II

Random factor effects model:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad (25.39)$$

where

- $\mu_{..}$ is a constant
- α_i, β_j and $(\alpha\beta)_{ij}$ are **independent** zero-mean normal random variables $N(0, \sigma_a^2)$, $N(0, \sigma_b^2)$, and $N(0, \sigma_{ab}^2)$
- ϵ_{ijk} are independent $N(0, \sigma^2)$.
- α_i, β_j , $(\alpha\beta)_{ij}$ and ϵ_{ijk} are pairwise independent
- $i = 1, ..., a$, $j = 1, ..., b$ and $k = 1, ..., n$.

Important Features:

1. Expected value of response: $Y_{ijk} = \mu_{..}$
2. Variance of response (total variance: components from A, B, A:B and errors):

$$var(Y_{ijk}) = \sigma_Y^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2$$

3. Observations have constant variance and are normally distributed. They are independent if they are from different levels, and correlated if they are from the same factor levels or treatment levels with the following covariance structure

- (a) from the same level i for factor A, but different levels for factor B (e.g. $A1B1$ vs. $A1B2$);
- (b) from the same level j for factor B, but different levels for factor A (e.g. $A1B1$ vs. $A2B1$);
- (c) from the same treatment level i, j for factors A & B (e.g. different observations from $A1B1$)
- (d) from different levels for factor A or factor B, (e.g. $A1B1$ vs. $A2B2$)

$$\sigma\{Y_{ijk}, Y_{ij'k'}\} = \sigma_\alpha^2 \qquad j \neq j' \qquad \textbf{(25.41a)}$$

$$\sigma\{Y_{ijk}, Y_{i'jk'}\} = \sigma_\beta^2 \qquad i \neq i' \qquad \textbf{(25.41b)}$$

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 \qquad k \neq k' \qquad \textbf{(25.41c)}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad i \neq i', j \neq j' \qquad \textbf{(25.41d)}$$

The correlation between two observations is the covariance divided by the total variance σ_Y^2.

Two-Factor Studies: ANOVA Model III

When one of the two factors has fixed factor levels while the other has random factor levels, a mixed factor effects ANOVA model (model III) is applicable.

The mixed ANOVA model for two-factor studies, where factor A is fixed and factor B is random as follows:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad (25.42)$$

where

- $\mu_{..}$ is a constant
- α_i are fixed (constant) effects subject to $\sum_i \alpha_i = 0$
- β_j are random effects , independent $N(0, \sigma_b^2)$.
- $(\alpha\beta)_{ij}$ are random and independent except they come from the same level of factor B.

$(\alpha\beta)_{ij}$ are $N\left(0, \dfrac{a-1}{a}\sigma^2_{\alpha\beta}\right)$, subject to the restrictions:

$$\sum_i (\alpha\beta)_{ij} = 0 \qquad \text{for all } j$$

$$\sigma\{(\alpha\beta)_{ij}, (\alpha\beta)_{i'j}\} = -\frac{1}{a}\sigma^2_{\alpha\beta} \qquad i \neq i'$$

- ϵ_{ijk} are independent $N(0, \sigma^2)$.
- β_j , $(\alpha\beta)_{ij}$ and ϵ_{ijk} are pairwise independent
- $i = 1, ..., a$, $j = 1, ..., b$ and $k = 1, ..., n$.

Important Features:

1. Expected value of response: $Y_{ijk} = \mu_{..} + \alpha_i$.
2. Variance of response (total variance: components from B, A:B and errors)

$$var(Y_{ijk}) = \sigma_Y^2 = \sigma_\beta^2 + \frac{a-1}{a}\sigma^2_{\alpha\beta} + \sigma^2$$

3. Observations have constant variance and are normally distributed. Observations are independent if they are not from the same level for factor B, and correlated if they are from the same factor level of factor B with the following covariance structure:

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\beta^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2 \qquad k \neq k' \qquad \textbf{(25.46a)}$$

$$\sigma\{Y_{ijk}, Y_{i'jk'}\} = \sigma_\beta^2 - \frac{1}{a}\sigma_{\alpha\beta}^2 \qquad i \neq i' \qquad \textbf{(25.46b)}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad j \neq j' \qquad \textbf{(25.46c)}$$

The correlation between two observations is the covariance divided by the total variance σ_Y^2.

Two factor ANOVA II and III Tables (balanced studies)

- The terms, layout, df of the ANOVA table are the same as what we used for the random or mixed effects model as compared to the standard ANOVA Table 1.

TABLE 19.8 **ANOVA Table for Two-Factor Study with Fixed Factor Levels.**

Source of Variation	SS	df	MS
Factor A	$SSA = nb\sum(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$a-1$	$MSA = \frac{SSA}{a-1}$
Factor B	$SSB = na\sum(\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$b-1$	$MSB = \frac{SSB}{b-1}$
AB interactions	$SSAB = n\sum\sum(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	$(a-1)(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$
Error	$SSE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij.})^2$	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$
Total	$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{...})^2$	$nab-1$	

- However, the expected mean squares (EMS) based on the ANOVA model I-III and their different random factor assumptions are different and resulting in different F-tests.

TABLE 25.5 **Expected Mean Squares for Balanced Two-Factor ANOVA Models.**

Mean Square	df	Fixed ANOVA Model (A and B fixed)	Random ANOVA Model (A and B random)	Mixed ANOVA Model (A fixed, B random)
MSA	$a-1$	$\sigma^2 + nb\frac{\sum\alpha_i^2}{a-1}$	$\sigma^2 + nb\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + nb\frac{\sum\alpha_i^2}{a-1} + n\sigma_{\alpha\beta}^2$
MSB	$b-1$	$\sigma^2 + na\frac{\sum\beta_j^2}{b-1}$	$\sigma^2 + na\sigma_\beta^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + na\sigma_\beta^2$
$MSAB$	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
MSE	$(n-1)ab$	σ^2	σ^2	σ^2

The correct F tests for model II-III are summarized in Table 25.6. Note we will not always use MSE as the denominator.

TABLE 25.6 Test Statistics for Balanced Two-Factor ANOVA Models.

Test for Presence of Effects of:	Fixed ANOVA Model (A and B fixed)	Random ANOVA Model (A and B random)	Mixed ANOVA Model (A fixed, B random)
Factor A	MSA/MSE	$MSA/MSAB$	$MSA/MSAB$
Factor B	MSB/MSE	$MSB/MSAB$	MSB/MSE
AB interactions	$MSAB/MSE$	$MSAB/MSE$	$MSAB/MSE$

Basically, each test statistic for testing certain factor effects is constructed by comparing two mean squares that have the properties:

1) Under H_0, both mean squares have the same expectation.
2) Under H_a, the numerator MS has a larger expectation than the denominator MS mean square. We need to compare the two MS so that numerator MS has an **extra** term than the denominator MS to describe the given factor to be tested. The two dfs for the F-test will use the dfs from the corresponding numerator and denominator MS.

Estimation and Inference

- For 2-factor random effects model, we are interested in estimating the overall mean $\mu_{..}$ and the variance component σ_a^2, σ_b^2, σ_{ab}^2 and σ^2, together with their standard errors for the estimators, and confidence intervals for the parameters.

- For 2-factor mixed effects model, e.g. assume factor A is fixed and B is random, then we are interested in estimating the overall mean $\mu_{..}$, α_i, and the variance component, σ_b^2, σ_{ab}^2 and σ^2, together with their standard errors for the estimators, and confidence intervals for the parameters.

Note: We will skip the Section 25.4 since the formulas are only applicable for balanced studies and the approximation method are not as efficient as the likelihood based methods to estimate the variance components and CI. We will illustrate next how to estimate and make inference for random and mixed models from maximum likelihood methods using **R**. (In **SAS**, PROC MIXED or PROC GLM with random statement may be used to get the MLE estimates.)

Example for ANOVA model II

Example : A manufacturer was developing a new spectrophotometer for medical labs. A critical issue is consistency of measurements from day to day among different machines. 4 machines were randomly selected from the production process and tested on 4 randomly selected days. Per day 8 serum samples were randomly assigned to the 4 machines (2 samples per machine). Response is the triglyceride level [mg/dl] of a sample. (From: Meier 2018. ANOVA: A Short Intro Using R)

1. We input the data into R

```
# read respsonse
y <- c(142.3, 144.0, 148.6, 146.9, 142.9, 147.4, 133.8, 133.2, ## day 1
       134.9, 146.3, 145.2, 146.3, 125.9, 127.6, 108.9, 107.5, ## day 2
       148.6, 156.5, 148.6, 153.1, 135.5, 138.9, 132.1, 149.7, ## day 3
       152.0, 151.4, 149.7, 152.0, 142.9, 142.3, 141.7, 141.2) ## day 4
```

```r
# add factor level to make a data frame
TRIG <- data.frame(Response = y, day = factor(rep(1:4, each = 8)),
                   machine = factor(rep(rep(1:4, each = 2), 2)))

str(TRIG)
```

```
## 'data.frame':    32 obs. of  3 variables:
##  $ Response: num  142 144 149 147 143 ...
##  $ day     : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 2 2 ...
##  $ machine : Factor w/ 4 levels "1","2","3","4": 1 1 2 2 3 3 4 4 1 1 ...
```

```r
c
```

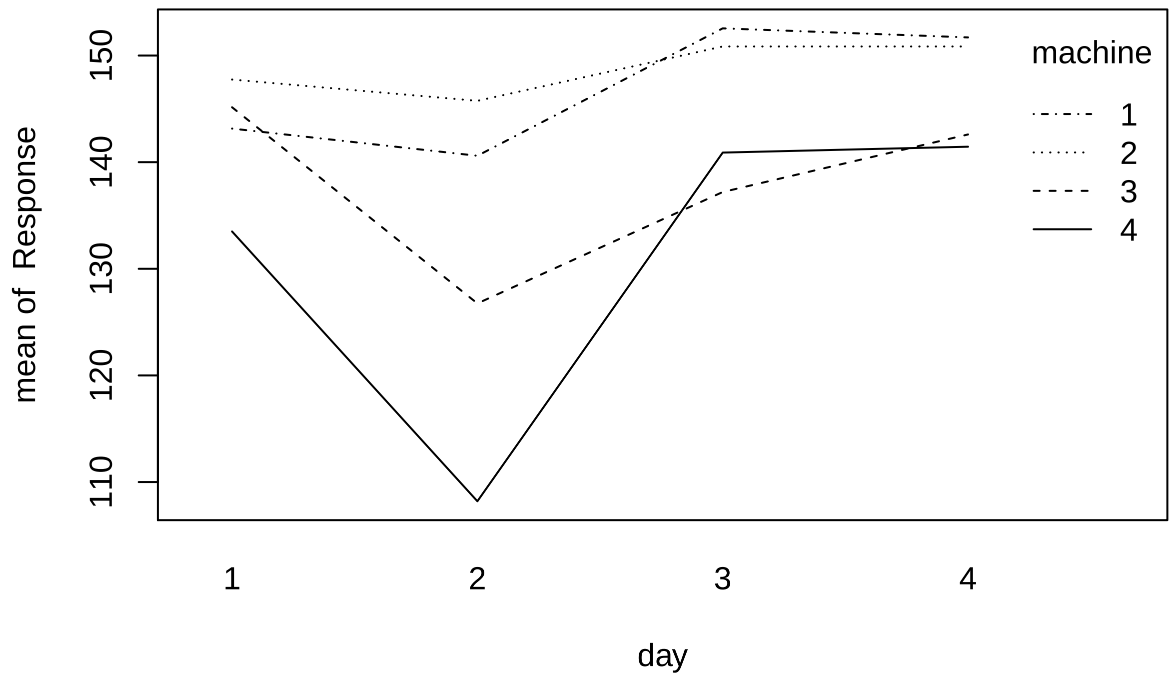
```
## function (...)  .Primitive("c")
```

```r
## verify number of observations
xtabs(~ day + machine, data = TRIG)
```

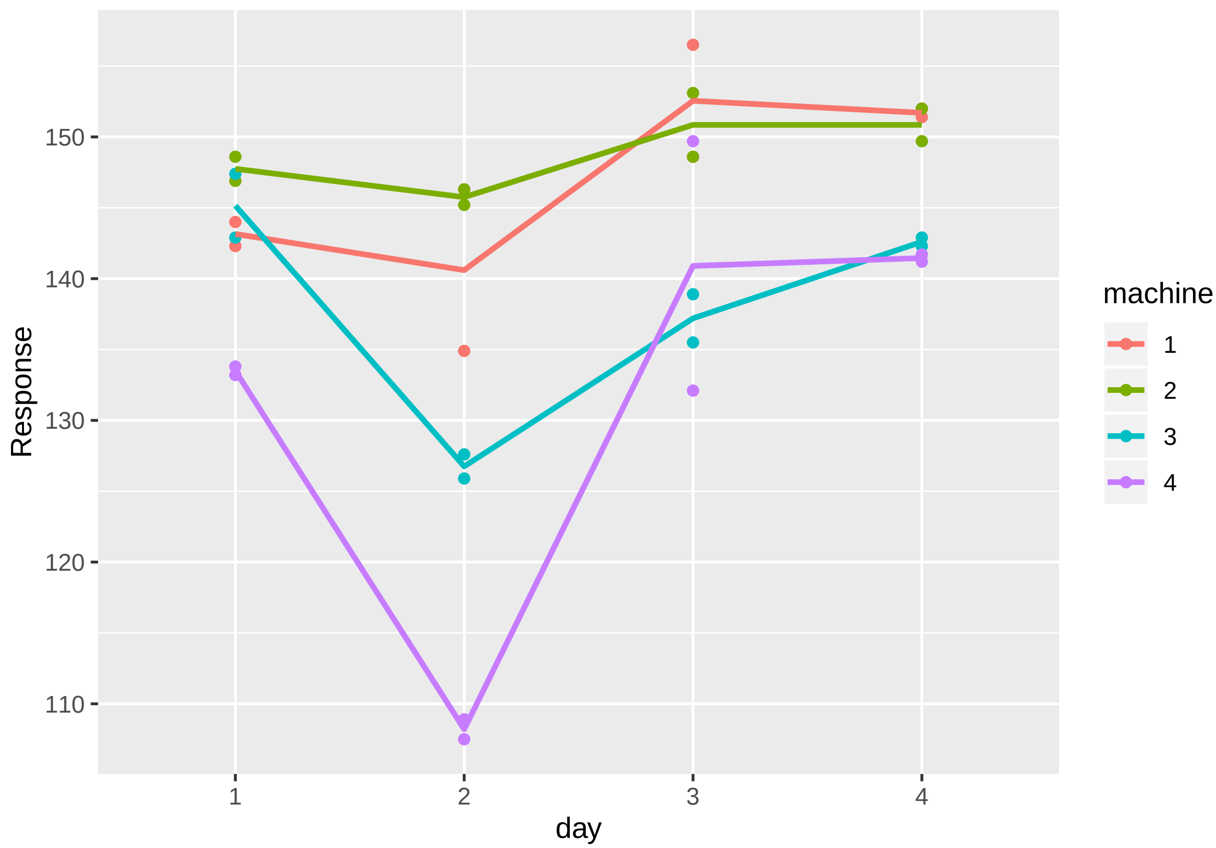
```
##    machine
## day 1 2 3 4
##   1 2 2 2 2
##   2 2 2 2 2
##   3 2 2 2 2
##   4 2 2 2 2
```

2. plot data and check interactions

```r
## approach 1:
with(TRIG, interaction.plot(x.factor = day, trace.factor = machine, response = Response))
```




```r
# use ggplot and group by machine
library(ggplot2)
ggplot(TRIG, aes(x =day , y = Response , group = machine , col = machine )) +
  geom_point() + stat_summary(fun.y = mean, geom = "line", size=1)
```



3. Testing: we need to generate standard ANOVA table, but will not use all standard F-tests and p-values

```
summary(aov(Response~day*machine, data= TRIG))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## day          3 1334.5   444.8   24.86 2.91e-06 ***
## machine      3 1647.3   549.1   30.68 7.19e-07 ***
## day:machine  9  786.0    87.3    4.88  0.00294 **
## Residuals   16  286.3    17.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: We can use all the MS, but we have to be careful to apply the correct F-test based on correct ANOVA model that we plan to use. For random effects model (ANOVA Model 2), from Table 25.5,

1. To test A:B interaction, we use $F^* = MSAB/MSE$, this is same as model I, therefore p-value in the above ANOVA table is correct, $p = 0.00294$, the model is not additive when looking at the effects of day and machine.

2. To test the main effects A and B, we use $F_A^* = MSA/MSAB \sim F(3, 9)$, and $F_B^* = MSB/MSAB \sim F(3, 9)$ and we can calculate the exact p-values.

```r
# For factor A
 (F.statA = 444.8/87.3)
```

```
## [1] 5.095074
```

```r
# p-value for testing factor A
 (pf.A= 1- pf(F.statA, 3, 9))
```

```
## [1] 0.02477665
```

```r
# For factor B
 (F.statB = 549.1/87.3)
```

```
## [1] 6.289805
```

```r
# p-value for testing factor B
 (pf.B= 1- pf(F.statB, 3, 9))
```

```
## [1] 0.01370686
```

Therefore, both the main effects were also significant at a significance level of 0.05. The response varies by the day and machine and their effects were not additive. It seems some day (day 2) and certain machine had the worse response compared to other day/other machines. (This may only suggest the researchers need to check the experiment setting/conditions for those day/machine combination to rule out data/human errors or abnormal testing samples or assays. They might need to run a large experiment if the day by day variation or the interaction is not expected.)

4. Estimate the overall mean and random components by ML (maximum likelihood) methods

We use (1 | day) + (1 | machine) + (1 | machine:day) for three random components α_i,β_j and $(\alpha\beta)_{ij}$ in the **lmer** function of **lme4** package.

```r
library(lme4)
```

```
## Loading required package: Matrix
```

```r
fit.TRIG <- lmer(Response ~ (1 | day) + (1 | machine) + (1 | machine:day), data = TRIG )
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.002216
## (tol = 0.002, component 1)
```

```r
summary(fit.TRIG)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Response ~ (1 | day) + (1 | machine) + (1 | machine:day)
##    Data: TRIG
```

```
##
## REML criterion at convergence: 215
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8426 -0.3558  0.0348  0.2072  2.3178
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  machine:day (Intercept) 34.69    5.890
##  machine     (Intercept) 57.77    7.601
##  day         (Intercept) 44.76    6.690
##  Residual                17.90    4.230
## Number of obs: 32, groups:  machine:day, 16; machine, 4; day, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  141.184      5.325   26.51
## convergence code: 0
## Model failed to converge with max|grad| = 0.002216 (tol = 0.002, component 1)
```

Results: From the random effect model output, we can get the estimated $\hat{\mu}_{..} = 141.184$ with $SE = 5.325$. The variance components $\hat{\sigma}_a^2 = 44.76$, $\hat{\sigma}_b^2 = 57.77$, $\hat{\sigma}_{ab}^2 = 34.69$ and $\sigma^2 = 17.90$. (The Std.Dev. column is just the square root of the variance, not the estimation error.)

Then we can also get the confidence interval for σ_{ab}, σ_b, σ_a, σ and overall mean $\mu_{..}$.

```r
confint(fit.TRIG, oldNames = FALSE)
```

```
## Computing profile confidence intervals ...
```

```
##                                 2.5 %     97.5 %
## sd_(Intercept)|machine:day   2.956021  11.296840
## sd_(Intercept)|machine       0.000000  18.481628
## sd_(Intercept)|day           0.000000  16.968920
## sigma                        3.100604   6.257749
## (Intercept)                129.938480 152.430210
```

Example for ANOVA model III (1)

Example 1: (p.1054) A study of four different training methods (factor A, fixed) and five instructors (factor B, random). Four classes were assigned to each training method–instructor combination. The response variable of interest was the mean improvement per student in the class at the end of the training program.

We have the following ANOVA table:

TABLE 25.7 **ANOVA Table for Mixed ANOVA Model—Training Example (A fixed, B random, $a = 4$, $b = 5$, $n = 4$).**

Source of Variation	SS	df	MS	F^*
Factor A (training methods, fixed)	42.1	3	14.0	$14.0/3.9 = 3.59$
Factor B (instructors, random)	53.9	4	13.5	$13.5/2.1 = 6.43$
AB interactions	46.7	12	3.9	$3.9/2.1 = 1.86$
Error	126.4	60	2.1	
Total	269.1	79		

Then we can set the correct F-test based on Model III, and calculate the p-value for the F-test. Note to use the corresponding df to get the critical values or compute the p-values.

```
# For interaction:
 (pf.AB= 1- pf(1.86, 12, 60))
```

```
## [1] 0.05840866
```

```
# For main effect A:
 (pf.A= 1- pf(3.59, 3, 12))
```

```
## [1] 0.04644538
```

```
# For main effect B:
 (pf.B= 1- pf(6.43, 4, 60))
```

```
## [1] 0.0002249746
```

Results: Therefore, for level of significance 0.05, the interaction effect (training: instructor) is not significant ($p = 0.06$). But we find that both training methods and instructors differ in effectiveness.

Example for ANOVA model III (2)

Example 2: We use the data set **Machines** from **nlme** package. In this study,data on an experiment to compare three brands of machines used in an industrial process. Six workers were chosen randomly among the employees of a factory to operate each machine three times. The response is an overall productivity score taking into account the number and quality of components produced. Here, the brands of machine is fixed, but the operator is a random factor.

1. We first load the data

```
# load the data and set the factors
data("Machines", package = "nlme")
Machines[, "Worker"] <- factor(Machines[, "Worker"], levels = 1:6, ordered = FALSE)

str(Machines, give.attr = FALSE) ## give.attr in order to shorten output
```

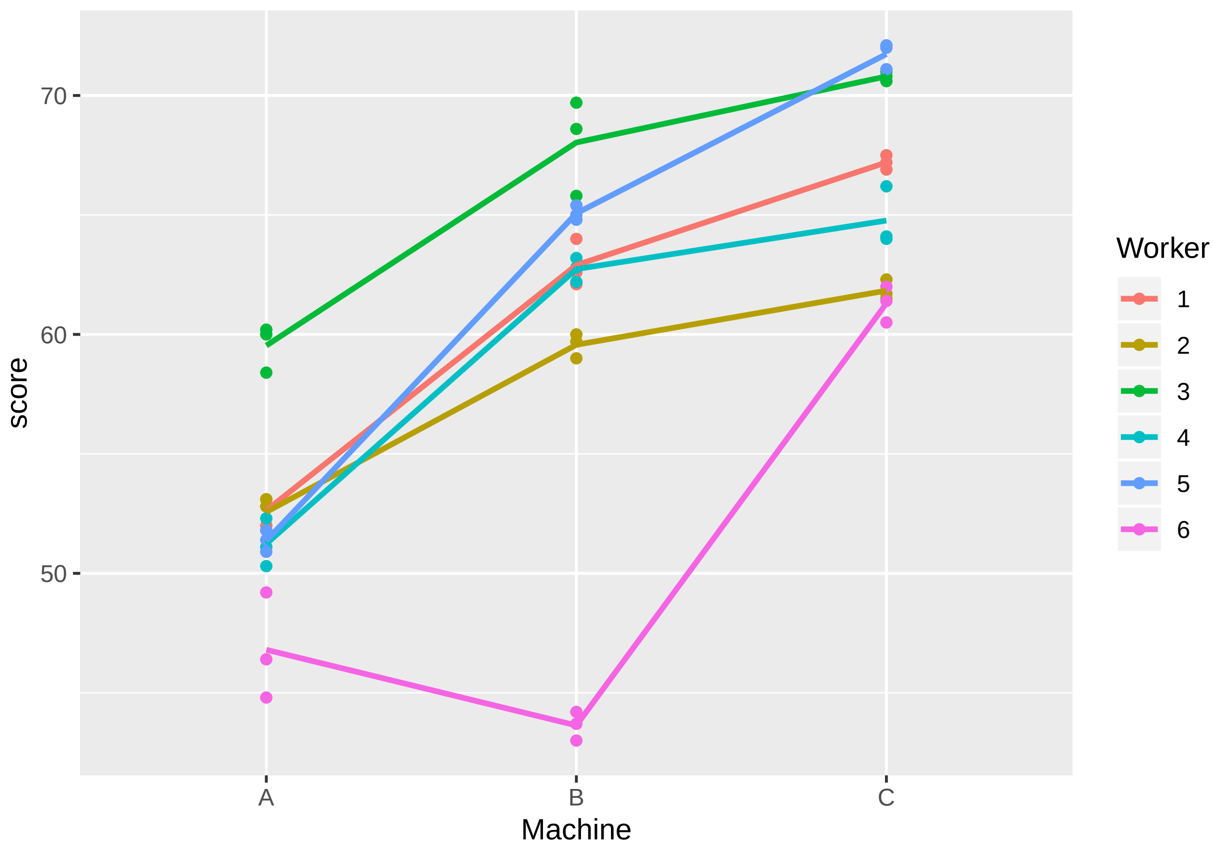
```
## Classes 'nffGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':   54 obs. of  3 variables
##  $ Worker : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ Machine: Factor w/ 3 levels "A","B","C": 1 1 1 1 1 1 1 1 1 1 ...
##  $ score  : num  52 52.8 53.1 51.8 52.8 53.1 60 60.2 58.4 51.1 ...
```

```
table(Machines$Machine)
```

```
##
##  A  B  C
## 18 18 18
```

2. Visualize the data

```
ggplot(Machines, aes(x = Machine, y = score, group = Worker, col = Worker)) +
  geom_point() + stat_summary(fun.y = mean, geom = "line", size=1)
```



Note: We observe that productivity is largest on machine C, followed by B and A. Most workers show a similar "profile", with the exception of worker 6 who performs badly on machine B.

3. We can perform the hypothesis testing based on ANOVA table.

```r
summary(aov(score  ~  Machine*Worker, data=Machines))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Machine         2 1755.3   877.6  949.17 <2e-16 ***
## Worker          5 1241.9   248.4  268.62 <2e-16 ***
## Machine:Worker 10  426.5    42.7   46.13 <2e-16 ***
## Residuals      36   33.3     0.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: Using similar steps as those for Example 1, we can calculate the correct F-test statistics and p-values. For interaction and random factor B, F-test is the same, we reject the null and there were significant interaction and main effect for worker. For the main effects of machine (fixed factor), we calculate the F-test with MSAB as denominator, then get the corresponding p-value for F(2, 10) distribution and it is also significant.

```r
# Testing stat and p-value for machine effects
(Fstat.A= 877.6/42.7)
```

```
## [1] 20.55269
```

```r
(pf.A= 1- pf(Fstat.A, 2, 10))
```

```
## [1] 0.0002868578
```

We note that although the machine effect is much more significant using fixed effect model. This is because the fixed effects model makes a statement about the machine effect of these 6 specific workers and not about the population average when we consider worker as a random factor.

4. Estimate the fixed and random components by ML (maximum likelihood) methods

- We set Machine as the fixed effect
- (1|Worker) for a random effect β_j per worker
- (1|Worker:Machine) for a random effect per combination of machine and workers
- We set the contrasts option to get the zero-sum constraints for fixed effects $\sum \alpha_i = 0$ (default is $\alpha_1 = 0$)

```r
options(contrasts = c("contr.sum", "contr.poly"))
fit.Machine <- lmer(score ~ Machine + (1 | Worker) + (1 | Worker:Machine), data = Machines)
summary(fit.Machine)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
##    Data: Machines
##
## REML criterion at convergence: 217.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.26959 -0.54847 -0.01071  0.43937  2.54006
```

```
##
## Random effects:
##  Groups         Name        Variance Std.Dev.
##  Worker:Machine (Intercept) 13.9095  3.7295
##  Worker         (Intercept) 22.8584  4.7811
##  Residual                    0.9246  0.9616
## Number of obs: 54, groups:  Worker:Machine, 18; Worker, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  59.6500     2.1447  27.813
## Machine1     -7.2944     1.2569  -5.804
## Machine2      0.6722     1.2569   0.535
##
## Correlation of Fixed Effects:
##          (Intr) Machn1
## Machine1  0.000
## Machine2  0.000 -0.500
```

Note: From the model output, *Random effects* section, we can get the estimated random variance components.

For fixed effects, we can estimate the overall mean $\hat{\mu}_{..} = 59.65$, and the main effects for for each machines $\hat{\alpha}_1 = -7.2944$ and $\hat{\alpha}_2 = -0.6722$, and the third machine would have $\hat{\alpha}_3 = -(\hat{\alpha}_1 + \hat{\alpha}_2) = 6.6222$. We can also get their confidence intervals easily.

```r
confint(fit.Machine, oldNames = FALSE)
```

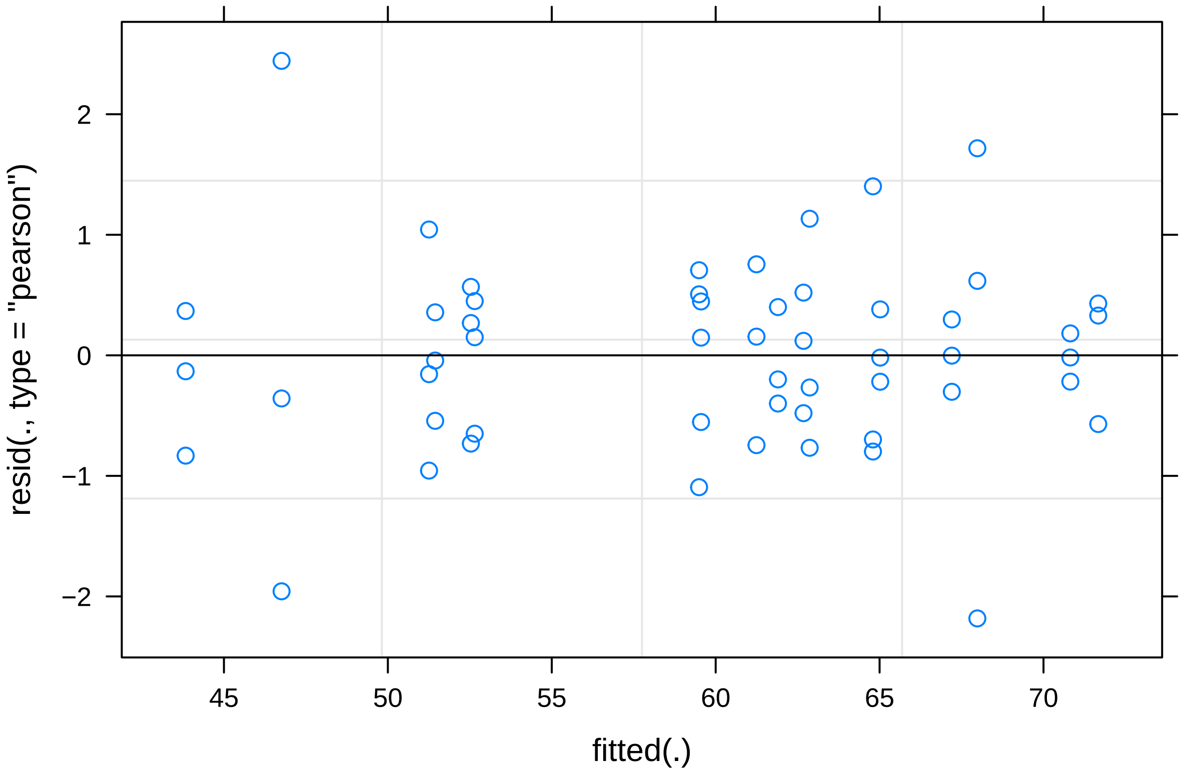
```
## Computing profile confidence intervals ...
```

```
##                                    2.5 %    97.5 %
## sd_(Intercept)|Worker:Machine  2.3528037  5.431503
## sd_(Intercept)|Worker          1.9514582  9.410577
## sigma                          0.7759506  1.234966
## (Intercept)                   55.1082569 64.191739
## Machine1                      -9.7358141 -4.853075
## Machine2                      -1.7691474  3.113592
```

Model checking for mixed effects model: Machines example

1. We can check the constancy of variance:

```r
plot(fit.Machine)
```



2. Check the normality of random effects and residuals

The **ranef()** function will provide the estimated random effects for the data.

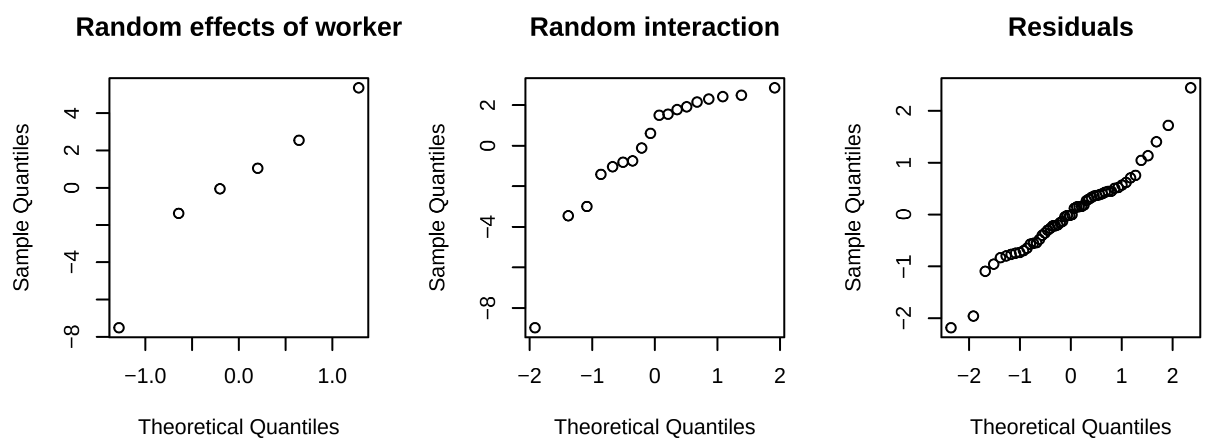
```
ranef(fit.Machine)
```

```
## $`Worker:Machine`
##     (Intercept)
## 1:A  -0.7501465
## 1:B   1.4999943
## 1:C  -0.1142371
## 2:A   1.5525649
## 2:B   0.6068536
## 2:C  -2.9966326
## 3:A   1.7776121
## 3:B   2.2993839
## 3:C  -0.8149413
## 4:A  -1.0393717
## 4:B   2.4173662
## 4:C  -1.4143951
## 5:A  -3.4569326
## 5:B   2.1521139
## 5:C   2.8532447
## 6:A   1.9162737
## 6:B  -8.9757118
```

```
## 6:C   2.4869615
##
## $Worker
##   (Intercept)
## 1  1.04454621
## 2 -1.37585602
## 3  5.36077680
## 4 -0.05981983
## 5  2.54464341
## 6 -7.51429057
##
## with conditional variances for "Worker:Machine" "Worker"
```

Then we can generate the QQ-plots for these random effects and residuals:

```r
par(mfrow = c(1, 3), pty='s')
qqnorm(ranef(fit.Machine)$Worker[, 1], main = "Random effects of worker")
qqnorm(ranef(fit.Machine)$'Worker:Machine'[, 1], main = "Random interaction")
qqnorm(resid(fit.Machine), main = "Residuals")
```



The QQ-plots look mostly OK as we don't have a lot of observations and these visual inspection is to spot the big departure from the model assumptions.

Summary

- Reading Chapter 25.1-25.4

- Homework (due next weekend, no class next Thursday)
  - 25.5, 25.6 (for d.- get 95% CI using the likelihood methods from software instead of Satterthwaite or MLS procedure) : one-factor random effect model and use data-set (CH25PR5-1factor-REM.txt).
  - 25.15 (for d and e: get 95% CI using the likelihood methods from software)- 2-factor random effects model and use data-set (CH25PR15-2factorREM.txt).
  - 25.16 (only need do questions (a-d), **skip (e-g)**)- 2-factor mixed effects model and use data-set (CH25PR16-2factorMEM.txt).
  - Datasets in class website (Week-13 folder).