Single-Factor Experiments: Analysis of Variance (ANOVA)

Marco Antonio Florenzano Mollinetti[1]

[1]**University of Tsukuba, Systems Optimization Laboratory**

mollinetti@syou.cs.tsukuba.ac.jp

Agenda

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Introduction

- Many single-factor experiments require more than two levels of the factor to be considered

- Randomization of the experiments is now taken into account

- In the medical field, effects of medicines/treatments are verified for multiple samples of populations

- We will now call each level a treatment

Introduction

■ Montgomery explains the steps for a experiment:

1. **Conjecture**: the original hypothesis that motivates the experiment.

2. **Experiment**: the test performed to investigate the conjecture.

3. **Analysis**: the statistical analysis of the data from the experiment.

4. **Conclusion**: what has been learned about the original conjecture from the experiment. Often the experiment will lead to a revised conjecture, and a new experiment, and so forth.

Introduction

- Factor levels can be chosen in two ways:
  - Fixed-effects model
  - Random effects/ Components of variance model

Introduction

- Factor levels can be chosen in two ways:
  - Fixed-effects model
    - Specifically choose the a levels
    - Conclusions cannot be extended to treatments that were not considered
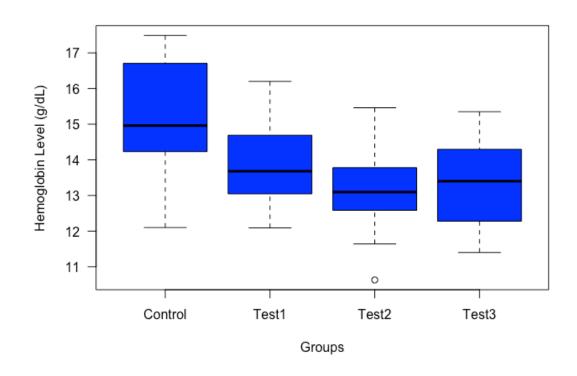 - Random effects/ Components of variance model

Introduction

- Factor levels can be chosen in two ways:
  - Fixed-effects model
  - Random effects/ Components of variance model
    - Random sample from a larger population of treatments
    - Extend the conclusion to all treatments
    - Knowledge about the treatments investigated is not important

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Completely Randomized single-factor Model

- Consider the following experiment: "We need to verify the effect of some medicine to anemic male patients. We have 4 groups: control, placebo and two variations of the medicine. Nominal hemoglobin levels for male are between 13.5 to 17.5g/dl. Sample size is 20. Age is disregarded

Completely Randomized single-factor Model

- Blood samples are taken in a completely randomized fashion

- Need to reduce any nuisance variable in the experiment

- Human error x Machine error

- Graphical interpretation and statistic interpretation

Completely Randomized single-factor Model

- Let's observe the boxplot of our data

Completely Randomized single-factor Model

- Suppose we have a different levels of a single factor we wish to compare

- Observations follow the linear model:

$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,n \end{cases}$$

Where $\mu_i = \mu + \tau_i$

Completely Randomized single-factor Model

- Suppose we have a different levels of a single factor we wish to compare
- Observations follow the linear model:

Mean of the treatment

Random observation

$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

Error

Where $\mu_i = \mu + \tau_i$

Effect factor

Mean of the population (overall mean)

Completely Randomized single-factor Model

- Each treatment can be thought of as a normal population with mean μ_i and variance σ^2

- Moreover, each treatment can be understood as a normally distributed variable with mean μ plus a perturbation τ_i

Completely Randomized single-factor Model

- For the fixed effect models, effects τ_i are considered as deviations from overall mean μ
- So the sum of τ_i is expected to be 0
- This is equivalent to testing the following hypotheses

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \text{ for at least one } i$$

Completely Randomized single-factor Model

- ANOVA partitions the variability into two parts
- Test the hypothesis based on a comparison of two independent estimates of the population variance
- We will consider the one-way ANOVA

Completely Randomized single-factor Model

- We have the following point estimators:

  - Total sum of squares: $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{\bar{y}^2}{N}$

  - Sum square of treatments: $SS_{Tr} = \sum_{i=1}^{a} \frac{y_i^2}{n} - \frac{\bar{y}^2}{N}$

  - Mean square of treatments: $MS_{Tr} = \frac{SS_{Tr}}{(a-1)}$

  - Error sum of squares: $SS_E = SS_T - SS_{Tr}$

  - Error mean square: $MS_E = \frac{SS_E}{a(n-1)}$

  - F-statistic: $F_0 = \frac{MS_{Tr}}{MS_E}$

Completely Randomized single-factor Model

■ Now let's run the ANOVA for our hemoglobin experiment

Since P is considerably smaller than 0.1 we can safely reject H_0

```
> summary(aov.ex1)
             Df  Sum Sq  Mean Sq  F value   Pr(>F)
m.hemo$Group  3    50.8   16.935    10.97  4.57e-06 ***
Residuals    76   117.3    1.543
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Completely Randomized single-factor Model
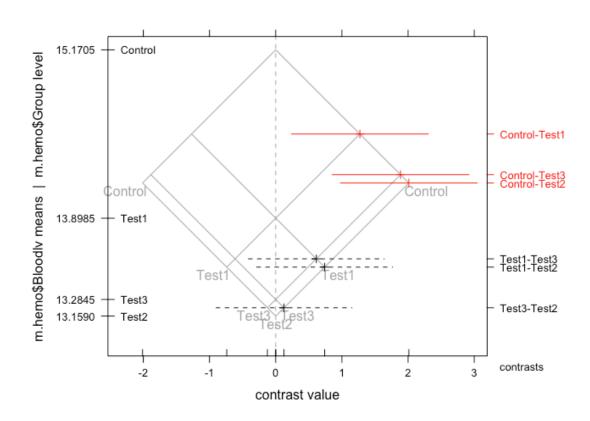
- Since the Mean Squared error MS_E is an estimator of variance σ, we can build confidence intervals
- Let's verify the confidence interval for each mean μ_i

Completely Randomized single-factor Model

- Since we rejected H_0, we know that there is at least one factor that is different from the others

- How do we know what factor?
  - ☐ Fisher's least significant Difference (LSD)
  - ☐ Tukey's test
  - ☐ mmc

- We'll run a mean-mean multiple comparison (mmc)

Completely Randomized single-factor Model

- Tukey test
  - Post-hoc analysis of the ANOVA
  - H_0: no significance in the difference of means
  - H_1: significance in the difference of means

Completely Randomized single-factor Model

- Plotting the mmc we get a very straightforward answer:

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Unbalanced Design

- In some cases, the number of observations taken under each treatment may be different
- We say the the design is unbalanced
- Disadvantages over balanced design:
  - Insensitive to small departures from the assumption of equality
  - More prone to Type-II errors, less power β

Unbalanced Design

- Load the assimetric_hemo_exp dataset
- Run the ANOVA
- Run the mmc
- Verify the differences

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Random Effects Model

- In many cases, the factor of interest has too many levels

- We want to draw conclusions for the entire population of factor levels

- a random factors are chosen

Random Effects Model

- Consider variance of the treatment effects τ_i to be σ_τ^2

- The variance of the response is $\sigma_\tau^2 + \sigma^2$

- So we test Hypotheses about σ_τ^2

$$H_0: \sigma_\tau^2 = 0$$
$$H_1: \sigma_\tau^2 \neq 0$$

Random Effects Model

- The computational procedure of the ANOVA table is the same as of the fixed-effects

- Conclusions, however apply for the entire population of treatments

- Load the 'Block_Hemo_exp.csv' dataset

Random Effects Model

- We will use a slightly different function rather than aov()
- Check the confidence intervals
- No mmc this time

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Blocking Design

- Reduce the variability from a nuisance factor
- Extension of the paired t-test when more than two treatments must be compared
- Selection of b blocks and running a complete replicate of the experiment in each block
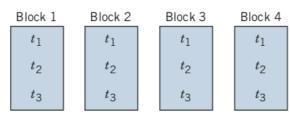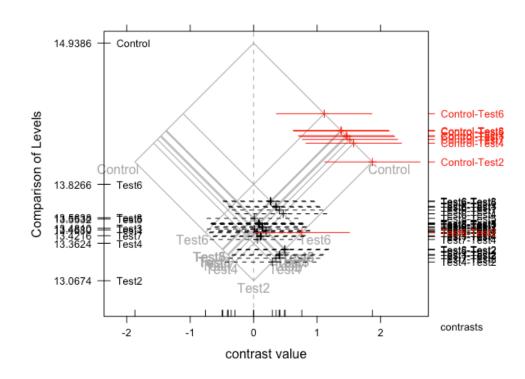- A levels, b blocks

Block 1	Block 2	Block 3	Block 4
t_1	t_1	t_1	t_1
t_2	t_2	t_2	t_2
t_3	t_3	t_3	t_3

Figure 13-8   A randomized complete block design.

Blocking Design

- When to use block design?
    - When you want to reduce the MS_E
    - When doing a single factor experiment has much more degrees of freedom
- Generally, it is based on trial and error

Blocking Design

- Run the ANOVA with block design at R
- Do the mmc
- Check the plot

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Model Validation

- Like the t-test, we must test our model for:
    - Normality
    - Independence
    - Heteroscedascity

Model Validation

- **Like the t-test**, we must test our model for:
    - □ Normality
    - □ Independence
    - □ Heteroscedascity
- **Unlike the t-test**, all of them are **based on plots:**
    - □ Normal Probability plot of residuals
    - □ Plotting the residuals against time
- Fortunately, we can obtain that by simply **using plot()** **on our ANOVA model!**

- Introduction
- Completely Randomized single-factor model
- Unbalanced Design
- Random Effects model
- Blocking design
- Model Validation
- Two way ANOVA

Two-way ANOVA

- Suppose now we want to analyze the effect of two factors

- Two-way ANOVA is the answer

- Hypothesis is still:

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$$
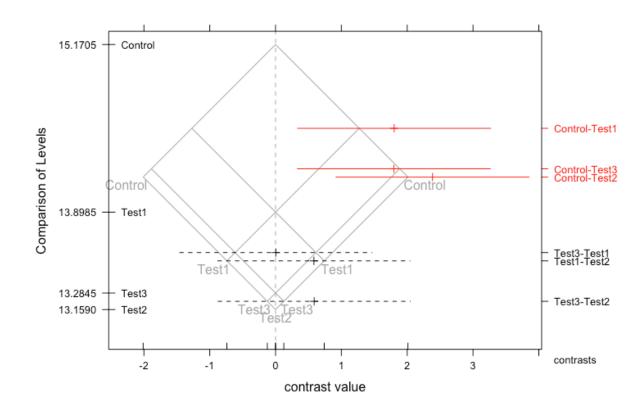$$H_1: \tau_i \neq 0 \text{ for at least one } i$$

Two-way ANOVA

- In R, we just add another factor to the equation in the aov()
- Same procedure as before
- Mmc and tukey test is also conducted

Two-way ANOVA

- Let's take a look at the mmc plot:

Next Episode

- ANOVA is not over yet! We will talk about ANOVA with multiple factors
- Non parametric tests for "non punchable" data

Single-Factor Experiments: Analysis of Variance (ANOVA)

Marco Antonio Florenzano Mollinetti[1]

[1]**University of Tsukuba, Systems Optimization Laboratory**

mollinetti@syou.cs.tsukuba.ac.jp