Designs to Study Variances I

Sean Hellingman ©

Design for Data Science (ADSC2030)

shellingman@tru.ca

Winter 2025

THOMPSON RIVERS UNIVERSITY

Topics

Introduction

- So far, we have covered only designs to analyse the changes in the response between different levels of controllable factors.
- Another purpose of experimentation is to study the sources of variability in the response.
- Understanding where the variability comes from, allows for more focused designs.

Illustrative Example 1

- Diagnostic tests conducted by medical professionals are known to vary with the procedures and equipment used.

- Experiments may be conducted to find out how much of this variation is due to procedure.
  - $\sigma_T^2 = \sigma_p^2 + \sigma_e^2$
    σ_T^2 is the total variance
    σ_p^2 is the portion due to the procedure
    σ_e^2 is the portion due to the equipment

- σ_p^2 and σ_e^2 are called the **components of variance/variance components.**

Motivations

1. Descriptive: The variance components have value in themselves.
2. Variance reduction: Gain insights into how to reduce the variance in the response.
3. Causes of variability: Identify the causes of variability that can be tested further in additional experiments.

Road Map

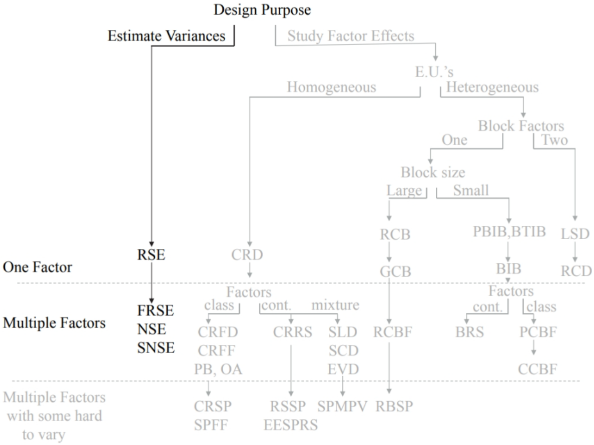

Figure: Source: (1)

Fixed Factors

1. When the purpose is to examine differences in the average response caused by differences in factor levels, the factors in the experiment are called **fixed factors**.
2. The *fixed factor levels* are selected by the experimenter.
3. In the effects model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, τ_i would be considered a fixed factor.

Random Factors

- In the effects model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, ϵ_{ij} would be considered a **random factor**.
- ϵ_{ij} represents the effect of the j^{th} experimental unit within the i^{th} level of the treatment factor.
- The levels of the random factors are just samples of the possible levels that could have been used.

Illustrative Example 2

- Recall in the bread dough example, there were $r = 4$ replicates of loaves used to estimate σ^2.
  - For each of the chosen *fixed* factor levels.
- The four replicate loaves used for each rise time represent only a sample of the loaves that could have been used in the experiment.
  - *Random samples*
- Experiments that are used to study variances can be thought of as **random sampling experiments** (RSE) since the factor levels are just a sample of the possible levels.

Model

- Change in the notation of models used to examine the variance:

$$y_{ij} = \mu + t_i + \epsilon_{ij}$$

- Where t_i represents the random effect of the treatment **when there is no interest in measuring the differences across the different levels.**

- ϵ_{ij} represents the effect of the j^{th} experimental unit within the i^{th} level of the treatment factor.

- Assumptions:
  - t_i and ϵ_{ij} are independent and normally distributed with zero means and variances equal to the variance components σ^2 and σ_t^2 respectively.

One-Factor Sampling Designs

- **One-factor sampling experiments** can be used to partition variability into two sources.
    - $\sigma_y^2 = \sigma_t^2 + \sigma^2$
- The partitioned variability can then be calculated and used to draw conclusions or with the design of further experiments.

Illustrative Example 3 I

- Recall the paper helicopter dropping experiment:
  - Response: The time it takes for a dropped paper helicopter to reach the ground.
- Suppose the definition of the experimental unit in this experiment is up for debate:
  - The sheet of paper used to make the helicopter.
  - The specific trial.
  - Air conditions when the helicopter was dropped.
- Partition the variability in drop times into helicopter-to-helicopter variability and variability among repeat drops of the same helicopter.

Illustrative Example 3 II

- If a substantial amount of the variability is due to helicopters of the same design then individual helicopters should be the experimental units.
- If the variability was due to drop times of the same helicopter, there would be no reason to make multiple helicopters of the same design for replicates.
  - Could drop the same helicopter multiple times as replicates.
- Cut and fold one standard helicopter with body width = 4.25, tail length = 4.0, and wing length = 6.5.
- Drop and time each of the six helicopters three times each according to a randomized order.

Estimation

- We will use the MLE or restricted maximum likelihood methods for estimation.
- Using this methodology prevents negative estimates of σ_t^2.
- We will use the *lme4* R package to arrive at numerical solutions for the unbalanced and balanced cases.

Estimation in R

- To estimate the variance components of one-factor sampling designs in R:
  - `library(lme4)`
  - `model <- lmer(response ~ 1 + (1|group), data = data)`
  - `summary(model)`

- In the summary, the $\hat{\sigma}_t^2$ is variance given to the `group` (variable name) and the $\hat{\sigma}^2$ is listed beside the `Residual`.

Example 1 Preliminaries I

- A manufacturer of packaged dry soup mixes was experiencing excessive variability in the **package weights** of a dry soup mix component called the *intermix*.
  - When incorrectly mixed, the flavour of the soup changes drastically.
- There are two steps to making the dried soup packets:
  1. Make a large batch of soup and then dry it.
  2. The *intermix* is then added to the dried soup in a large mixer.
- There are several factors at each step that could be causing the weight variations.

Example 1 Preliminaries II

- In order to determine which factors to include in a factorial experiment:
  1. Partition the variability in *intermix* weight into the variability from soup batch to soup batch and the variability within a batch caused by the process to mix and add *intermix* to the dried soup.
  2. If there is little to no variability from batch to batch, only need to consider factors involved in the mixing step.

Example 1

- Import the *soupmx* data into R.
- Take a moment to understand the data.
- Use the `lmer` function to partition the variance.
- What do these results imply?

Confidence Intervals

- An asymptotic approximate confidence interval for the variance components can be obtained in the R:
  - `pr1 <-profile(fm1M <-lmer(response ~ 1 + (1| group), data = data, REML = FALSE))`
  - `confint(pr1)`
- `.sigma` gives the approximate confidence interval for σ
- `.sig01` gives the approximate confidence interval for σ_t

Example 2

- Calculate an asymptotic approximate confidence interval for the variance components of the experiment in Example 1.
- Do these results surprise you?

Sample Size

- Important things to consider when estimating two variance components:
  1. To accurately estimate the replicate variance (σ^2), the number of degrees of freedom for the error ($\nu_2 = t(r-1)$) is very important.
  2. The accuracy of the estimate of σ_t^2 will always be relative to the accuracy of the estimate of σ^2.

- The expected width of the 95% confidence interval:

$$\sigma^2 \left[\frac{t(r-1) \cdot \left(\chi^2_{t(r-1),0.975} - \chi^2_{t(r-1),0.025} \right)}{\chi^2_{t(r-1),0.975} \cdot \chi^2_{t(r-1),0.025}} \right] \tag{1}$$

Calculating Sample Size I

- Based on the expected width of the confidence interval one can calculate a desired t and r.
- To search for the values of t and r such that the multiplier in (1) is 1.0 using R:
  - `nu2 <-36:44`
  - `chiu <- qchisq(.975, nu2)`
  - `chil <- qchisq(.025, nu2)`
  - `width <- nu2 * (chiu - chil) / (chil * chiu)`
  - `halfw <- width/2`
  - `data.frame(nu2, width, halfw)`

Example 3

- Run the following code to determine how many degrees of freedom for the error ν_2 is needed such that the multiplayer in (1) is 1.0:
  - `nu2 <-36:44`
  - `chiu <- qchisq(.975, nu2)`
  - `chil <- qchisq(.025, nu2)`
  - `width <- nu2 * (chiu - chil) / (chil * chiu)`
  - `halfw <- width/2`
  - `data.frame(nu2, width, halfw)`

Rule of Thumb

- When σ_t^2 is expected to be larger than σ^2, t should be as large as possible.
- Reasonable allocation: $t = \nu_2$ and $r = 2$.

Calculating Sample Size II

- Another way to determine t and r would be to consider the power of the F-test used to test: $H_0 : \sigma_t^2 = 0$ & $H_1 : \sigma_t^2 > 0$.

$$1 - \beta = P\left(F_{t-1,t(r-1)} > \frac{1}{1 + r + \rho} \cdot F_{t-1,t(r-1),\alpha}\right) \tag{2}$$

- where $\rho = \sigma_t^2/\sigma^2$

Example 4

- Run the following code to determine several combinations of r and t to have a power of a test $(1 - \beta)$ greater than 0.9 for a $\rho = \sigma_t^2/\sigma^2 \geq 3.0$.
  - `alpha <-.05`
  - `rho <-3.0`
  - `t <-rep(5:7, each = 3)`
  - `r <-rep(2:4, 3)`
  - `nu_1 <-t-1`
  - `nu_2 <-t * (r - 1)`
  - `fcrit <-qf( 1 - alpha, nu_1, nu_2 )`
  - `factor <-1 / ( 1 + r * rho )`
  - `plimit <-factor * fcrit`
  - `power <-1 - pf( plimit, nu_1, nu_2 )`
  - `data.frame( t, r, power)`

Calculating Sample Size Comments

- The second method does not consider the accuracy of the estimate of σ^2.
- Recommendation:
  1. Use the first method (I) to determine ν_2 to have a desired accuracy of σ^2.
  2. Use the second method (II) to determine how large t should be conditioned on ν_2 (already calculated) to achieve a desired power for the test when $\sigma_t^2/\sigma^2 \geq \rho$.

Two-Factor Sampling Designs

- Designs are similar to the two-factor factorial designs we have covered when the purpose is to study the variance in the response caused by varying the levels of two independent factors.
- In the **two-factor sampling experiment** the factor levels are *random* or represent a sample of possible levels.
  - Determine how much of the variability in the response is caused by the factors.
- Called **factorial random sampling experiments** (FRSE).

Road Map

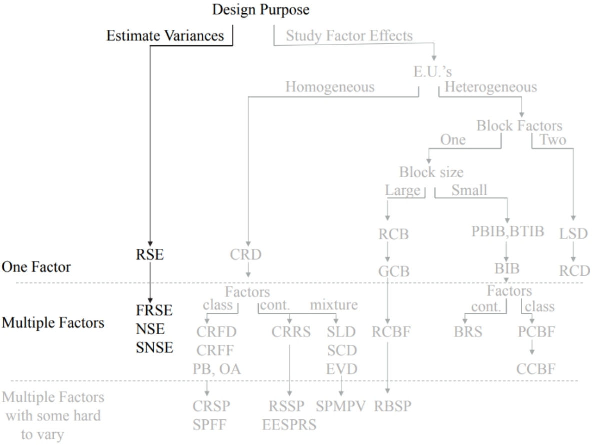

Figure: Source: (1)

Two-Factor Sampling Experiment Model

- Model:

$$y_{ijk} = \mu + a_i + b_j + ab_{ij} + \epsilon_{ijk}$$

- The effects a_i, b_j, and ab_{ij} are assumed to be normally distributed random variables with zero means and variances σ_a^2, σ_b^2, and σ_{ab}^2.
- Variability of the response: $\sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma^2$

Two-Factor Sampling Experiment in R

- Luckily, we can use R to estimate these variance components:
  - `library(lme4)`
  - `model <- lmer(response ~ 1 + (1|factor.A) + (1|factor.B) + (1|factor.A:factor.B), data = data)`
  - `summary(model)`
- The variability caused by each component can be examined.
- *We will use the next example to discuss interpretation.*

Example 5 Preliminaries I

- *The purpose is to classify the variability measured in features of manufactured products or product components.*

- **Gauge repeatability**: A single operator's ability to obtain the same measurement multiple times using the same measuring instrument on the same part.

- **Gauge reproducibility**: The ability of *different* operators to obtain the same measured value multiple times using the same gauge on the same part.

- *If the variability in measurements caused by the gauge repeatability plus the gauge reproducibility is more than 10% of the tolerance range, the measurements may not be accurate enough to be used in monitoring product quality.*

Example 5 Preliminaries II

- Data:
  - Part-to-part variability: A sample of 10 parts represents the levels in the first factor.
  - Inspector variability: A random or representative sample of inspectors represents the levels in the second factor.
  - Response: Each inspector randomly measures each part two times.
- Model: $y_{ijk} = \mu + a_i + b_j + ab_{ij} + \epsilon_{ijk}$
- y_{ijk} is the k^{th} measurement made by the j^{th} operator on the i^{th} part.
- a_i is the part effect, b_j is the inspector effect, and ab_{ij} is the interaction.

Example 5 Preliminaries III

- It is now assumed that a_i, b_j, and ab_{ij} are independent normally distributed random variable.
  - With zero means and variances: σ_a^2, σ_b^2, and σ_{ab}^2
- Total variance in the response (because independent):
  - $\text{Var}(y) = \sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma^2$
    - σ_a^2 caused by the actual differences in part features.
    - σ_b^2 caused by differences in inspectors.
    - σ_{ab}^2 caused by the interaction of the inspector and part.
    - σ^2 caused by the replicate measurements or gauge repeatability.
    - $\sigma_b^2 + \sigma_{ab}^2$ gauge reproducibility.
    - $\sigma_b^2 + \sigma_{ab}^2 + \sigma^2$ measure of the variance attributable to measurement error (repeatability plus reproducibility).

Example 5

- Import the *gagerr* dataset from the *daewr* package into R to complete the following tasks:
  1. Examine the ANOVA table for the model shown on Example 5 Preliminaries II.
  2. Use the `lmer()` function to calculate the variance components.
  3. What percentage of the measurement error is due to reproducibility (inspector accuracy)?
     - $\sigma_{ab}^2/(\sigma_b^2 + \sigma_{ab}^2 + \sigma^2)$
  4. What do these results indicate?

Confidence Intervals

- An approximate confidence interval for the variance components of two-factor designs can be obtained in the R:
  - `library(daewr)`
  - `vci(confl, c1, ms1, nu1, c2, ms2, nu2)`
    - `confl = 1 -` α
    - `c1 = c1` *positive confidence value*
    - `ms1 = MS1` *mean squared from ANOVA*
    - `nu1 = df1` *degrees of freedom*
    - `c2 = c2` *positive confidence value*
    - `ms2 = MS2` *mean squared from ANOVA*
    - `nu2 = df2` *degrees of freedom*

Example 6

- Using the ANOVA table from Example 5, estimate the 90% confidence interval for the variance component of the inspectors (σ_b^2).
  - Assume `c1 = c2 = 0.05`

- *Note: the estimation method differs*

Sample Sizes for Two-Factor Studies

- Three sample sizes need to be considered:
  - Number of levels of factor A (a)
  - Number of levels of factor B (b)
  - The number of replicates within each cell (r)

- Replace $t(r-1)$ with $ab(r-1)$ in Equation (1).

- Use the methods already covered to determine the value of $ab(r-1)$ needed for the desired interval width.

Unequal Replication

- The estimation method used by *lme4* package in R allows for computation of variance components even with unequal replication.
- The ANOVA table **cannot** be used to estimate confidence intervals.
- By performing an ANOVA on the cell means approximate confidence intervals may be estimated.

Cell Means

- The first step is to obtain the cell means and the ANOVA table.
- In R:
  - `cellmeans <-tapply( data$response, list(data$factor.A, data$factor.B), mean)`
  - `dim(cellmeans) <- NULL`
  - `factor.A <- factor(rep(c(1,2,...,a), each = b) )`
  - `factor.B <- factor(rep(c(1,2,...,b), a) )`
  - `model <-aov(cellmeans ~ factor.A + factor.B + factor.A:factor.B)`
  - `summary(model)`

Unequal Replication Confidence Intervals

- The same methodology used to estimate confidence intervals for two-factor sampling designs may be applied.
  - `library(daewr)`
  - `vci(confl, c1, ms1, nu1, c2, ms2, nu2)`

    - `confl = 1 -` α
    - `c1 = c1` *from equation below*
    - `ms1 = MS1` *mean squared from ANOVA*
    - `nu1 = df1` *degrees of freedom*
    - `c2 = c2` *from equation below*
    - `ms2 = MS2` *mean squared from ANOVA*
    - `nu2 = df2` *degrees of freedom*

$$\bar{c} = ab/\sum_i \sum_j (1/r_{ij})$$

$$c_1 = c_2 = \frac{1}{b\bar{c}}$$

Example 7 Preliminaries

- *Results from a sampling study to estimate the sources of variability in an inter-laboratory assay of calcium in blood serum.*
  - lab: (Laboratory) factor with three levels (A, B, C)
  - sol: (Standard Solution) factor with four levels (1, 2, 3, 4)
  - calcium: (Response) calcium in blood serum
- *There is unequal replication throughout*

Example 7

- Import the *blood* dataset and conduct the following analysis:
  1. Use the `lmer()` function to examine the variance components of the study.
  2. Use the cell means to estimate the confidence intervals for the variance of the laboratory (σ^2_{lab}).

Exercise 1

- A manufacturer of sports drinks was experiencing excessive variability in the *concentration* of the electrolytes component. Using the *SportsDrink.csv* data conduct the following analysis:
    1. Use the `lmer()` function to partition the variability.
        - What do these results imply?
    2. Calculate an asymptotic approximate confidence interval for the variance components of the experiment.

Exercise 1

- A manufacturer of sports drinks was experiencing excessive variability in the *concentration* of the electrolytes component. Using the *SportsDrink.csv* data conduct the following analysis:
    1. Use the `lmer()` function to partition the variability.
        - What do these results imply?
    2. Calculate an asymptotic approximate confidence interval for the variance components of the experiment.

Exercise 2

- The manufacturing company in Example 5 changed their training regime and re-conducted the entire experiment two years later.
- Using the *gagerr2.csv* dataset repeat Example 5 and Example 6.
    - Are the implications of these results any different?

References & Resources

1. Lawson, J. (2014). *Design and Analysis of Experiments with R (Vol. 115)*. CRC press.

- lme4 package
- `lmer()`
- `profile()`
- `vci()`