

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DEPARTMENT OF
INDUSTRIAL ENGINEERING 

Design of Experiments

Lesson #4 – Flipped lesson

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: <https://research.dii.unipd.it/capelab/>

Flipped lesson

- Today's lecture will be a **flipped classroom**
 - the learning procedure is inverted
- Please, complete the following procedure before Lesson #21:
 1. **attend this video lecture**
 2. **self-assess your learning**
 3. **read the following book chapter**
 - Montgomery, D. (2013). Design and Analysis of Experiments. *J. Wiley & Sons.*
 - Chapter 3
 4. **prepare questions** and anything you need to discuss with the teacher and your mates in the next lecture
- Next lecture will be held in the following manner:
 - 1/2 Q&A:
 - questions (of the students) and answers (of the teacher)
 - 1/2 example

Analysis of variance

Different treatments in a single factor

- Suppose a different **treatments/levels** of a **single factor** should be compared
- The n observed response from each of the a treatments is a **random variable**
- The data are organized in matrices \mathbf{Y} whose elements y_{ij} represent the j -th observation taken under factor level or treatment i
 - n observations are available for the i -th treatment
 - $i \in [1, a]$
 - $j \in [1, n]$

Modelling data: means model

- The observations from an experiment can be effectively described with a **model**

- One way to write this model is the **means model**:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

- where μ_i is the **mean** of the i -th factor treatment/level
- ε_{ij} is a **random error** component that incorporates all other sources of variability in the experiment:
 - measurement error
 - variability arising from uncontrolled factors
 - differences between the experimental units (such as test material, etc...)
 - general background noise in the process
- It is convenient to think of the errors as having mean zero

Modelling data: effect model

- An alternative way to write a model for the data is to define:

$$\mu_i = \mu + \tau_i$$

- where μ the **overall mean** that is a parameter common to all treatments
- τ_i is a parameter unique to the i -th treatment called the **i -th treatment effect**

- In this form the model is an **effects model**:

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

- Both the means model and the effects model are **linear statistical models**
 - response variable y_{ij} is a linear function of the model parameters
 - the effects model is more widely encountered in the literature
 - it has an intuitive appeal in that:
 - μ is a constant
 - the treatment effects τ_i represent deviations from this constant when the specific treatments are applied

One-way (single factor) analysis of variance

- Both the means model and the effects model are called **one-way analysis of variance (ANOVA)** model (or **single-factor ANOVA**):
 - only one factor is investigated
 - requirement: the experiments are performed in random order
 - the environment in which the treatments are applied (often called the **experimental units**) is kept as **uniform** as possible
 - the experimental design is a **completely randomized design**
- **Objectives of ANOVA:**
 - to test appropriate hypotheses about the treatment means
 - to estimate them
- For hypothesis testing, the **model errors** are assumed to be *normally and independently distributed random variables* with mean 0 and variance σ^2
 - σ^2 is assumed to be constant for all levels of the factor
 - this implies that the observations $y_{ij} = N(\mu + \tau_i, \sigma^2)$ and that the observations are mutually independent

Fixed and random factors

- The ANOVA model describes two different situations with respect to the treatment effects:
 1. **fixed effects model**: the a treatments could have been specifically chosen by the experimenter
 - we test hypotheses about the treatment means
 - conclusions apply only to the factor levels considered in the analysis
 - cannot be extended to similar treatments that were not explicitly considered
 - the model parameters can be estimated
 2. **random effects model** (or components of variance model): the a treatments could be a *random sample* from a larger population of treatments (not treated in this course)
 - extend the conclusions which are based on the sample of treatments to all treatments in the population, whether or not explicitly considered in the analysis
 - the τ_i are random variables
 - knowledge about the particular ones investigated is relatively useless
 - test hypothesis about the variability of the τ_i
 - estimate this variability of the τ_i

Starting points for self-assessment

- What is ANOVA? What can be used for?
- Are there any assumptions on the basis of ANOVA?
- What is the formulation of a means model?
- What is the formulation of effects model?

Analysis of fixed effects

Treatment equality testing

- We are interested in **testing the equality of the a treatment** means
 - the appropriate hypotheses are:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

$$H_1: \mu_i \neq \mu_j \quad \text{for at least one pair } (i, j)$$

- In the effects model, we break the i -th treatment mean μ_i into two components $\mu + \tau_i$

- where μ is the overall mean

$$\frac{\sum_{i=1}^a \mu_i}{a} = \mu$$

- this implies that the treatment can be thought of as deviations from the overall mean

$$\sum_{i=1}^a \tau_i = 0$$

- Testing the equality of treatment means means testing that the treatment effects are zero

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1: \tau_i \neq 0 \quad \text{for at least one } i$$

Analysis of variance

- The **analysis of variance** is carried out **partitioning the total variability in its components**
- The total variability is determined by:
 - $y_{i.}$: total of the observations under the i -th treatment
 - $\bar{y}_{i.}$: average of the observations under the i -th treatment
 - $y_{..}$: grand total of all the observations
 - $\bar{y}_{..}$: grand average of all the observations
 - $N = an$: total number of observations

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

total sum of squares

$$y_{i.} = \sum_{j=1}^n y_{ij}$$

$$\bar{y}_{i.} = y_{i.}/n$$

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^n y_{ij}$$

$$\bar{y}_{..} = y_{..}/N$$

Fundamental ANOVA identity

- The **fundamental ANOVA identity** is made of two contributions:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

- a sum of squares of the **differences between the treatment averages and the grand average**
 - a measure of the differences between treatment means
 - a sum of squares of the **differences of observations within treatments** from the treatment average
 - due to only random error
- The fundamental identity rewritten in symbolic manner is:

$$SS_T = SS_{\text{Treatments}} + SS_E$$

total s. s. = s. s. of the treatments + s. s. of the error

Terms of the fundamental ANOVA identity

- In summary, the ANOVA identity provides two estimates of the variance σ^2 :

- **variability between treatments:**

- if there were no differences between the a treatment means, we could use the variation of the treatment averages from the grand average to estimate the variance

$$\frac{SS_{\text{Treatments}}}{a - 1} = \frac{n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2}{a - 1}$$

- **variability within treatments:**

- the term within square brackets, if divided by $(n-1)$, is the sample variance in the i -th treatment
- the **pooled estimate** of the common variance within each of the a treatments is:

$$\frac{\sum_{i=1}^a \left[\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \right]}{\sum_{i=1}^a (n - 1)} = \frac{SS_E}{(N - a)}$$

- The variance σ^2 can be estimated by the **mean squares**:
 - the expected value of the mean square of the error is the variance σ^2

$$MS_E = \frac{SS_E}{N - a}$$

- the expected value of the mean square of the treatments is the variance σ^2 , as well, but only if the means are equal:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a - 1}$$



- Heuristically we can deduce that a **test of the hypothesis of no difference in treatment** means can be performed by comparing MS_E and $MS_{\text{Treatments}}$!

What could be an appropriate test to compare mean squares?

- Observations and errors can be assumed to be normally and independently distributed:
 - the total sum of squares SS_T are distributed as a chi-square distribution with $(N - 1)$ degrees of freedom
 - since the degrees of freedom for $SS_{\text{Treatments}}$ and SS_E add to $(N - 1)$, the total number of degrees of freedom, Cochran's theorem implies that $SS_{\text{Treatments}}$ and SS_E are independently distributed chi-square random variables with $(a - 1)$ and $(N - a)$ degrees of freedom, respectively
- The null hypothesis of **no difference in treatment means** is true if the ratio F_0 is distributed as a **F-distribution with $(a - 1)$ and $(N - a)$ degrees of freedom**:

$$F_0 = \frac{SS_{\text{Treatments}}/(a - 1)}{SS_E/(N - a)} = \frac{MS_{\text{Treatments}}}{MS_E}$$

Analysis of variance

- Under the null hypothesis:
 - MS_E is an unbiased estimator of σ^2
 - also $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2
- If the null hypothesis is false: $E(MS_{\text{Treatments}}) > \sigma^2$



$$F_0 > F_{\alpha, a-1, N-a}$$

- under the alternative hypothesis, H_0 should be rejected to conclude that there are differences in the treatment means
- the F test approximates a **randomization test** (computationally intensive)
- Alternatively, we could use the **p -value approach** for decision making

Estimator of the model parameters

- Being the error normally distributed, as well as the treatment means, a **confidence interval** can be defined for the treatment means using MS_E as variance estimator:

$$\bar{y}_i - t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_i + t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}}$$

- these are one-at-a-time intervals
- The confidence interval can be used to test if the difference in means are meaningful
- A total of r simultaneous confidence intervals on treatment means can be calculated with the **Bonferroni method**
 - the confidence interval is the same as the one showed before, but with a $\alpha/(2r)$ confidence limits

ANOVA model reliability

ANOVA model verification

1. Checking model adequacy

null hypothesis
failed to reject

null hypothesis
rejected

2. Understanding what are the treatment means which significantly differ when the null hypothesis is rejected

Comparison among treatments means

- Suppose that in conducting ANOVA for the fixed effects model the null hypothesis is rejected
 - there are differences between the treatment means
 - but exactly *which* means differ?
 - further comparisons and analysis **among groups of treatment means** may be useful
- The procedures for making these comparisons are usually called **multiple comparison methods**
 - comparisons between treatment means are made in terms of either:
 - treatment totals
 - treatment averages
 - See Matlab® command: **multcompare**
- Several methodologies for the comparison on treatments means are available (no details will be given in this course about them):
 - analysis of contrasts/orthogonal contrasts
 - contrasts are linear combinations among parameters
 - t-tests on contrasts or F-tests on their sum of squares are available
 - **Scheffè's method**
 - **Tuckey's test** on pair of means

$$H_0: \sum_{i=1}^a c_i \mu_i = 0$$

$$H_1: \sum_{i=1}^a c_i \mu_i \neq 0$$

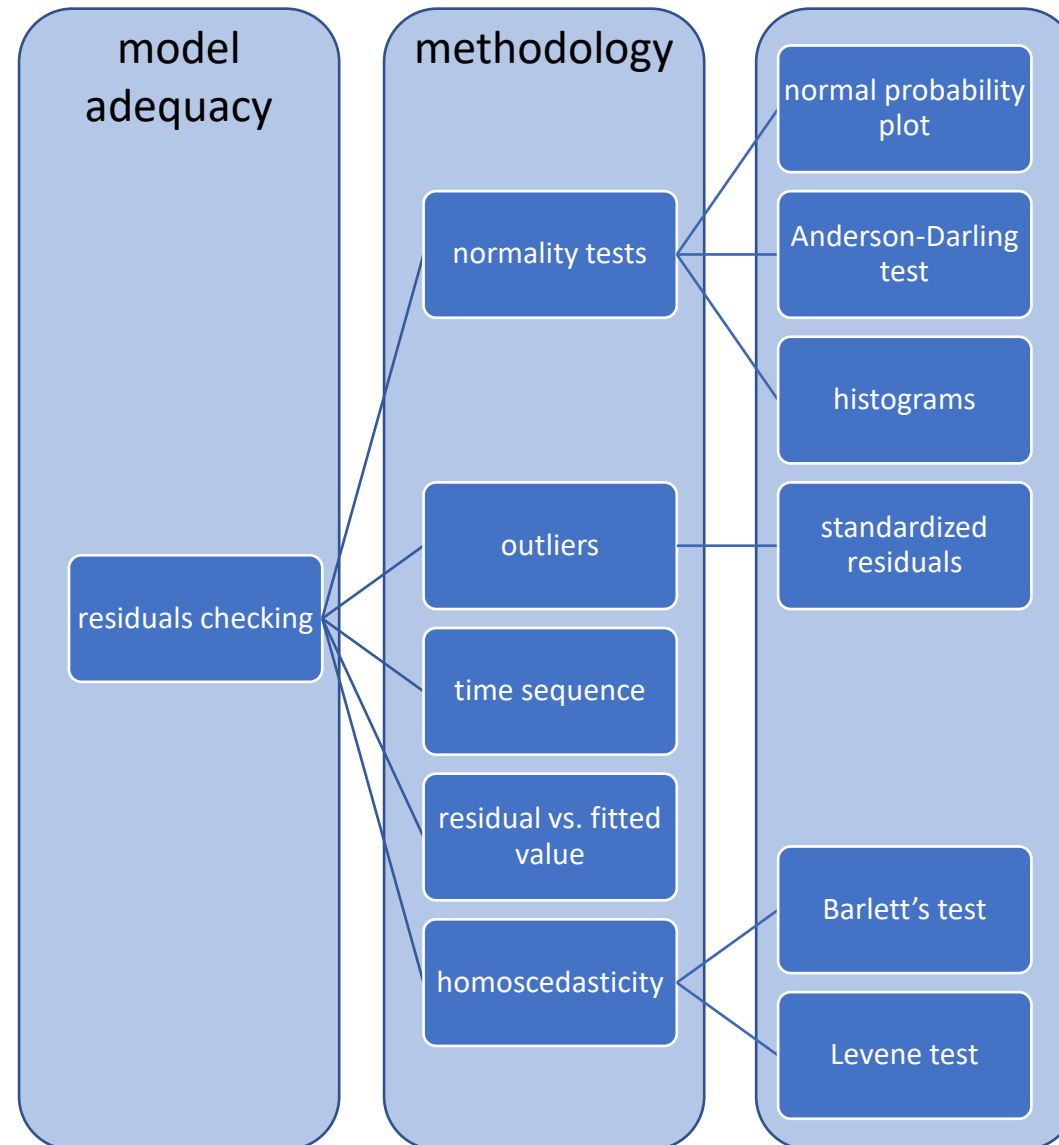
Model adequacy checking

- Analysis of variance is a reliable test only if the above-mentioned assumptions are satisfied
- The **errors** must be **normally and independently distributed**:

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

- violations of the basic assumptions and model adequacy can be easily investigated by the examination of **residual**
- **Examination of the residuals** is part of ANOVA:
 - if the model is adequate, the residuals should be **structureless**
 - many types of model inadequacies and violations of the underlying assumptions can be discovered through residuals

Residuals verification for model adequacy checking



Residuals normality assumption

- Methods to check the normality assumption for residuals:
 - **histogram of the residuals:**
 - if the $NID(0, \sigma^2)$ assumption on the errors is satisfied, this plot should look like a sample from a normal distribution centered at zero
 - with small samples, considerable fluctuation in the shape of a histogram often occurs
 - **normal probability plot of the residuals:**
 - effective tool when few observations are available for each treatment
 - if the underlying error distribution is normal, this plot will resemble a straight line
 - in visualizing the straight line, more emphasis on the central values of the plot than on the extremes should be given
- *Moderate departures from normality are of little concern* in the fixed effects analysis of variance:
 - an error distribution that has considerably thicker or thinner tails than the normal is of more concern than a skewed distribution
 - if one residual much larger than any of the others, an **outlier** is present: presence of one or more outliers can seriously distort the analysis of variance
 - careful investigate the motivation of the outlier
 - the **Anderson-Darling test** can be carried out to identify outliers

Standardized residuals

- Standardizing the residuals with the estimated variance MS_E , it is possible to observe the **standardized residuals**:

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}}$$

- being them $N(0,1)$:
 - ~67% of the observations should stay within the space among $[-1,1]$
 - ~95% of the observations should stay within the space among $[-2,2]$
 - 99.73% of the observations should stay within the space among $[-3,3]$

Residual time sequence

- Violation of the **independence assumption** is a potentially serious problem
 - difficult to correct
 - it is important to prevent the problem if possible
- Proper **randomization** of the experiment is an important step in obtaining independence
 - sometimes the skill of the experimenter may change as the experiment progresses
 - the process being studied may drift or become more erratic
 - this will often result in a change in the error variance over time
 - residuals versus time exhibit more spread in time
 - non-constant variance is a serious problem
- **Residuals in the time sequence plot:**
 - it is helpful to detect **correlation** between the residuals
 - a tendency to have runs of positive and negative residuals indicates positive correlation
 - this would imply that the **independence assumption** on the errors has been violated

Residuals versus fitted values

- Since the residuals should not display any structure/pattern, in particular, they should be unrelated to any other variable including the predicted response
- A simple check is to plot the **residuals versus the fitted values**:
 - for the single-factor experiment model, remember that for the i -th treatment:
$$\hat{y}_{ij} = \bar{y}_i.$$
 - this plot should not reveal any obvious pattern
- A defect that occasionally shows up on this plot is **non-constant variance**
 - sometimes the variance of the observations increases as the magnitude of the observation increases
 - this would be the case if the error or background noise in the experiment was a constant percentage of the size of the observation
 - this commonly happens with many measuring instruments
 - non-constant variance also arises in cases where the data follow a non-normal, skewed distribution because in skewed distributions the variance tends to be a function of the mean
 - if the assumption of homogeneity of variances is violated, the F test is only slightly affected
 - in unbalanced designs or in cases where one variance is very much larger than the others, the problem is more serious especially if the factor levels having the larger variances also have the smaller sample sizes
 - this is a good reason for choosing **equal sample sizes** whenever possible
 - variance could be stabilized with some nonlinear transformations
 - square root
 - logarithmic, etc...

Statistical tests for homoscedasticity

■ Bartlett's test:

- hypothesis testing:
- formulation:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

$$H_1: \text{above not true for at least one } \sigma_i^2$$

$$\chi_0^2 = 2.3026 \frac{q}{c}$$



$$q = (N - a) \log_{10} S_p^2 - \sum_{i=1}^a (n_i - 1) \log_{10} S_i^2$$
$$c = 1 + \frac{1}{3(a - 1)} \left(\sum_{i=1}^a (n_i - 1)^{-1} - (N - a)^{-1} \right)$$
$$S_p^2 = \frac{\sum_{i=1}^a (n_i - 1) S_i^2}{N - a}$$



$$\chi_0^2 > \chi_{\alpha, a-1}^2$$

- very sensitive to normality assumption

■ Modified Levene test:

- hypothesis testing: the same as in the previous case
- formulation: to test the hypothesis of equal variances in all treatments, the modified Levene test uses the absolute deviation of the observations in each treatment from the treatment median

$$d_{ij} = |y_{ij} - \tilde{y}_i| \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n_i \end{cases}$$

- robust to departures from normality

Starting points for self-assessment

- What is the fundamental ANOVA identity? What are their terms?
- What is the hypothesis testing formulation for ANOVA?
- How could you verify if you fail to reject the null hypothesis in the ANOVA model?
- How do you check the ANOVA model adequacy?

... per sempre a fianco a me!

