Seminar

Experimental Design and Hypothesis Testing

Ali Ashasi-Sorkhabi, PhD, EIT
Bryan Tolson, PhD, PEng.
Liping Fu, PhD, PEng.
Giovanni Cascante PhD, PEng.

September 4, 2019

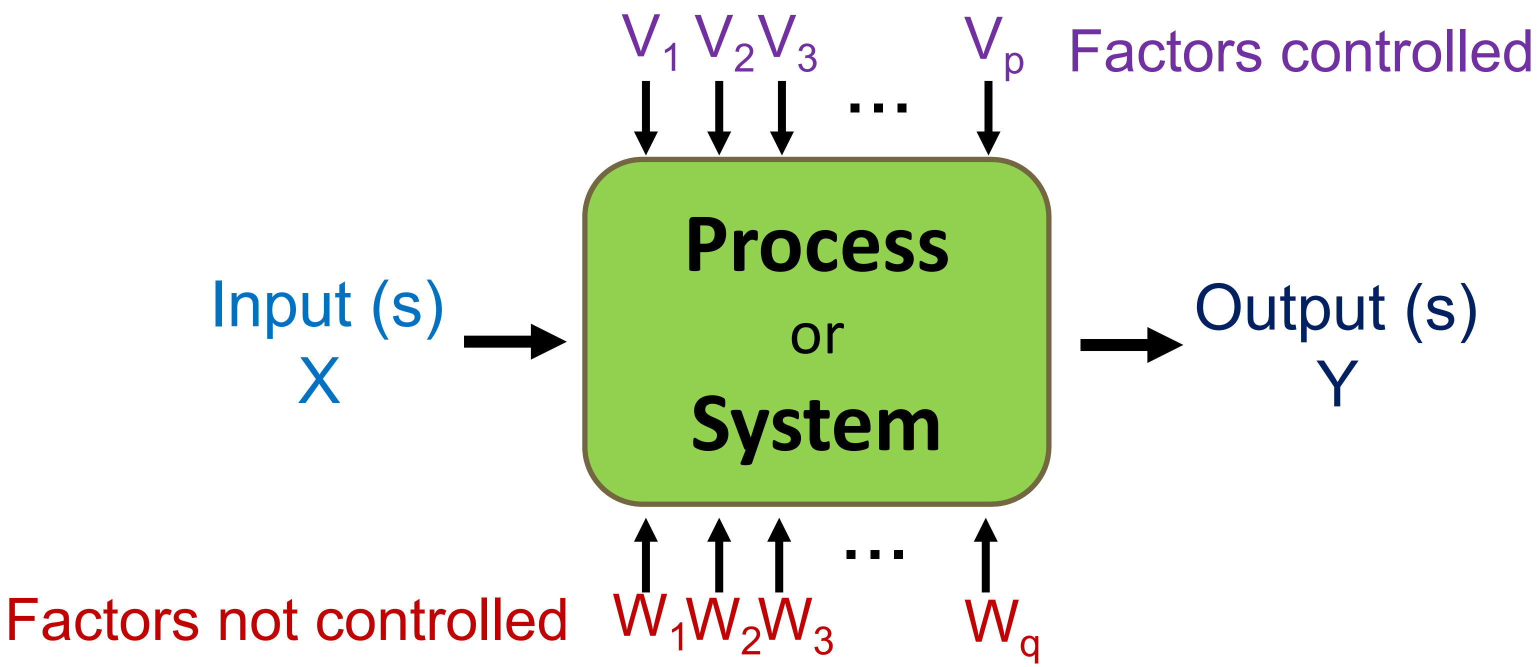
Experimental Design and Hypothesis Testing

Agenda

1. **Experimental design**
   - What is an experiment?
   - Strategy of experimentation
   - Design guidelines
2. **Simple comparative experiments**
   - Hypothesis testing
   - Confidence intervals
   - Choice of sample size
3. **Error analysis methods**
   - Time independent data
   - Time dependant data

What is an experiment?

- In general experiments are conducted to study the behaviour of a particular **process** or **system**; each experimental run is also called a test.
- General model of a process/ system



What is an experiment?

Process/system:

- o involves one or a combination of **operations**, **machines**, **methods**, **people** or other resources
- o transforms some **input, X,** (often a material) into an **output, Y,** that has one or more observable response variables.
- o includes **variables/factors** and **material properties** that can be **controlled** whereas other ones **cannot.**

What is an experiment?

Typical **objectives** of an experiment are to determine:

- which variables are most influential on the response Y
- the optimal values of V's so that the desired output, Y is obtained
- the values of V's so that variability in Y is small
- the values of the influential V's to minimize the effects of external factors, W's, are minimized.

⇨ **to develop a reliable model for the system/process**

What is an experiment?

Two types of models:

- **Empirical models: data driven models.**
  - Representative data is a "must" to build/verify model
  - Based on cause and effect between variables, not much else is needed in terms of process knowledge
- **Mechanistic models:** based on the underlying **physics** and **chemistry principals** affecting the behaviour of system
  - Use fundamental/physical knowledge of process
  - Conduct experiments to determine the parameters
  - Not much data needed for model development
  - More reliable

What is an experiment?

Remarks:

- Experiments involve several factors (controlled & not)
- Main goal is to determine the influence of these factors on the system response
- A successful experiment should result in a reliable model for the system

Strategy of experimentation

- The approach for planning and doing the experiment is called the **strategy of experimentation.**
- Questions about how to vary the factors affecting the system/process
- There are different strategies for experimentation:
    - Best-guess approach
    - One-factor-at-a-time (OFAT)
    - Factorial experiment

Strategy of experimentation

- Example: Playing Golf
  - Objective: improve the score
  - Possible influential parameters:
    - Type of the driver (regular/over sized)
    - Type of the ball (balata/3-piece)
    - Walking and carrying the golf clubs/ riding in a golf card
    - Drinking water/other while playing
    - Play time (morning/afternoon)
    - Playing in cool or hot weather
    - .....
  - Condition: maximum of **8 rounds** of golf can be played

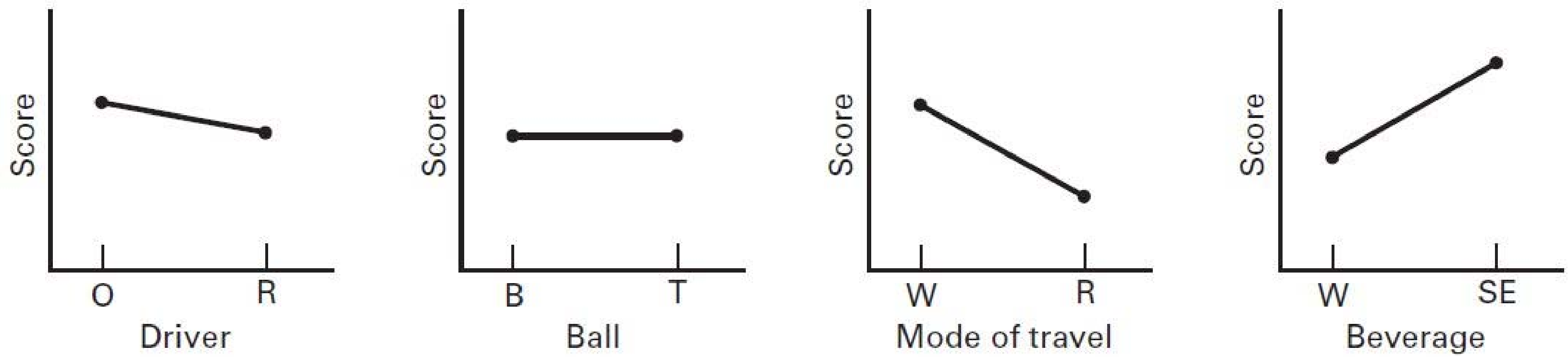
Strategy of experimentation

- **Best-guess approach:**
  - Frequently used in practice by engineers
  - Consider an arbitrary combination of factors and see what happens next:
    - Round 1: oversized driver-balata ball-golf card-water
    - Round 2: regular-sized driver-balata ball-golf card-water
    - .... This approach could be continued almost indefinitely !
  - **Pros:** Could work reasonably well!
  - **Cons:** if the 1st guess is not good, 2nd one needed, **long time!** Also, If the 1st guess gives acceptable results, experimenter may stop, but no guarantee that best solution was found

Experimental Design and Hypothesis Testing

Strategy of experimentation

- **One-factor-at-a-time (OFAT):**
  - Select a baseline of levels for each factor, then vary each one over its range with other factors held constant at baseline



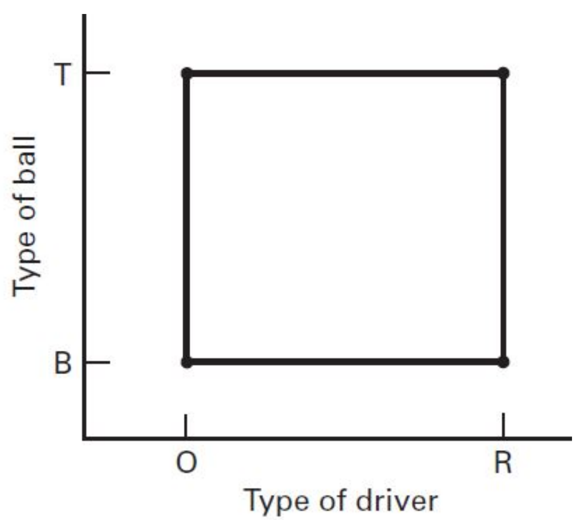
  - **Pros**: interpretation of OFAT graphs is straightforward.
  - **Cons**: it fails to consider the **interactions** between factors that are very common and if they occur, the OFAT produces **poor results**

Experimental Design and Hypothesis Testing

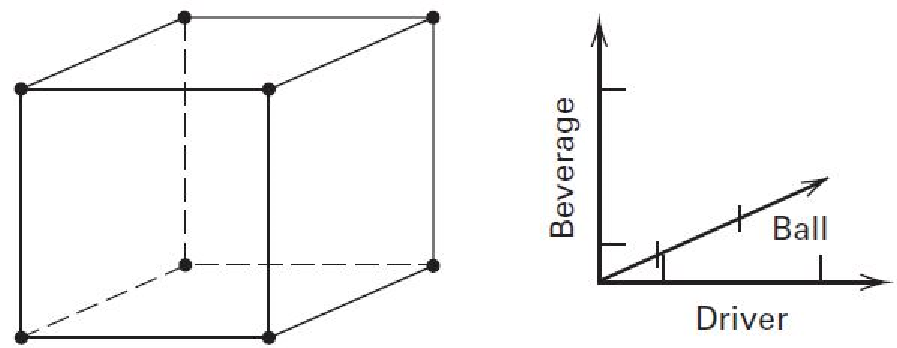
Strategy of experimentation

- **Factorial experiment:**
  - Factors are varied together rather than one at a time.

2^2 factorial design



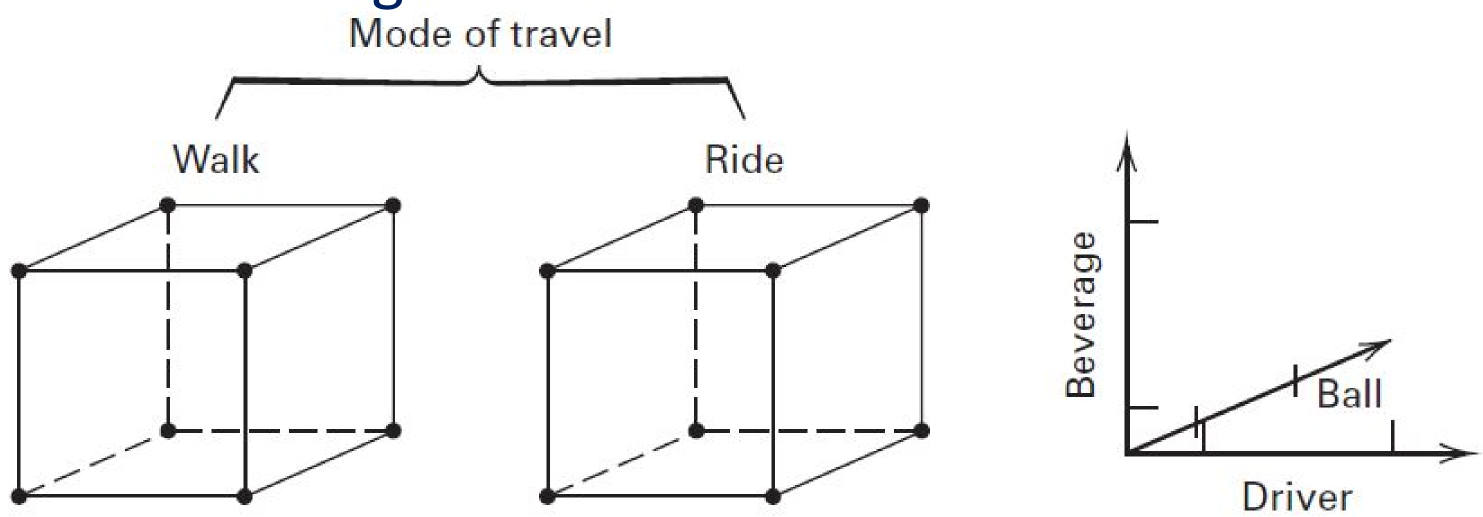
2^3 factorial design



Experimental Design and Hypothesis Testing

Strategy of experimentation

- 2^4 factorial design



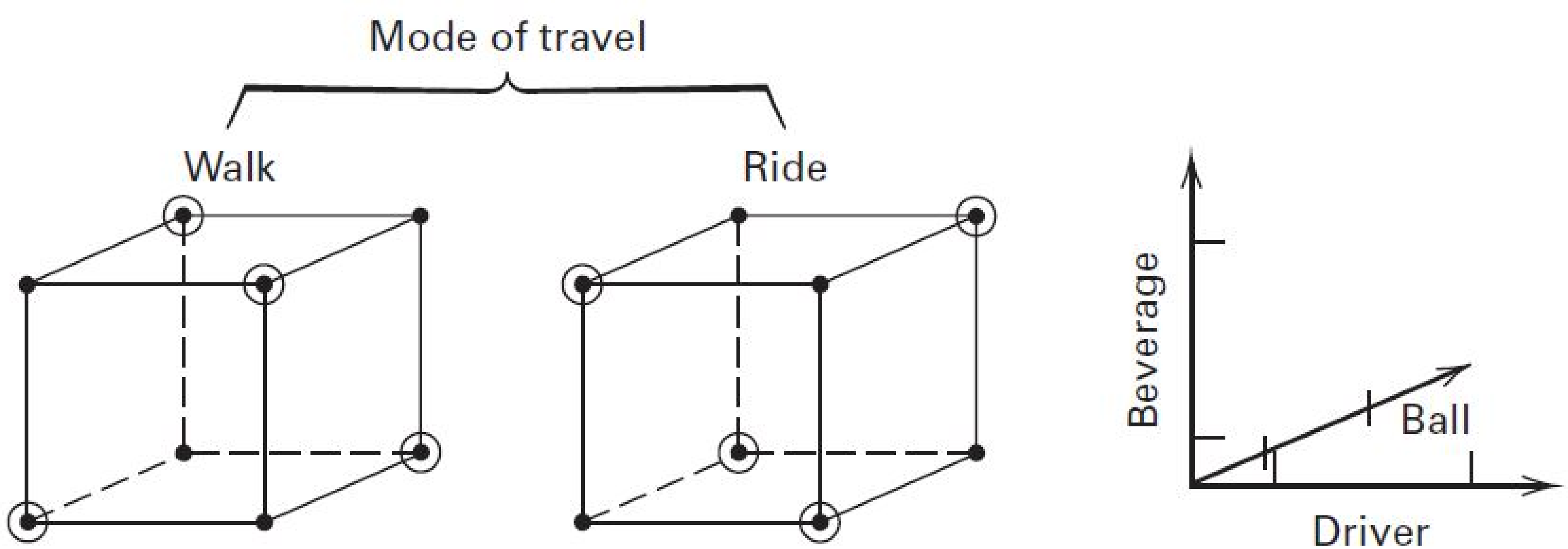
- **Factorial experiment:**
  - o If there are k factors, each at 2 levels, 2^k tests would be needed.
  - o If the number of factors of interest increases (e.g. k=10), the number of runs increases very fast (1024 runs). It becomes **impractical** from **time** and **resources** points of view

Experimental Design and Hypothesis Testing

Strategy of experimentation

- **Fractional factorial experiment:**

2^4 factorial design (one-half fraction)



General guidelines for designing an experiment

1. Recognition of of and statement of the problem
2. Selection of the response variables
3. Choice of factors, levels, and ranges

(Steps 1–3: Pre-experimental planning)

4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of results
7. Conclusions and recommendations

Steps 2 and 3 are often done together or with reversed order

Remarks:

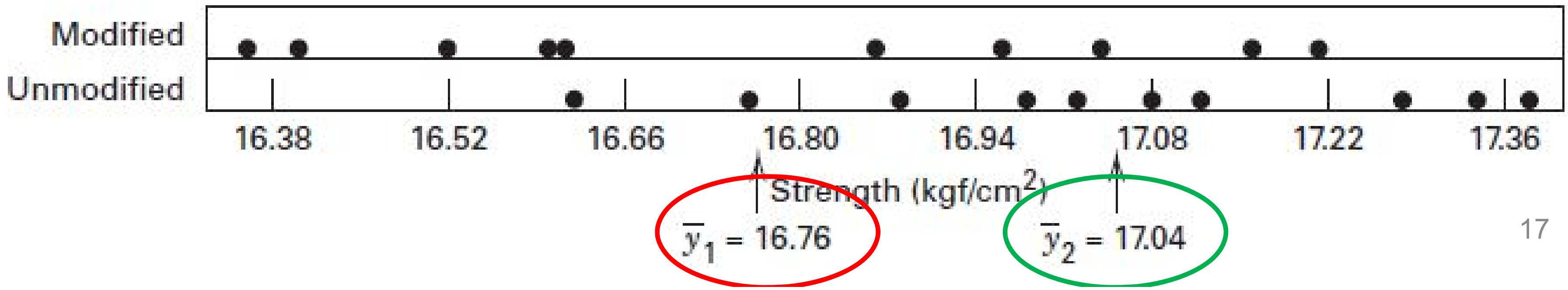
1. Use your non-statistical knowledge of the problem
2. Keep the design and analysis as simple as possible
3. Recognize practical and statistical significances
4. Experiments are usually iterative

Initial experiment < %25 of resources of testing (runs, time, money)

Simple comparative experiments

- Experiments to **compare two conditions**
- Example: **formulation of a Portland cement mortar**
  - o **Objective:** compare the results obtained from method 1 and 2
  - o 2 different formulations ~ **2 levels of the factor formulation**
  - o **Findings:**
    - ➢ Large reduction in the cure time
    - ➢ Bond strength:

visual average comparison



Experimental Design and Hypothesis Testing

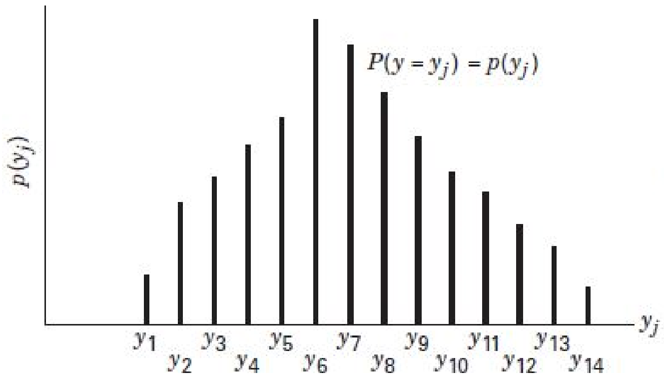
Simple comparative experiments

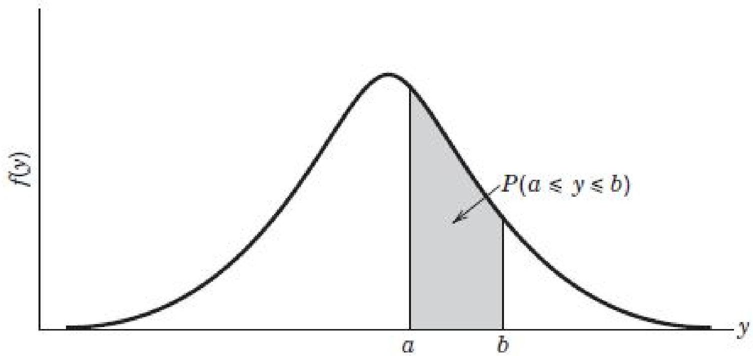
- **Hypothesis testing**
  - Allows the comparison of the two formulations in objective terms, knowing the risks associated with a wrong conclusion.

- **Review of basic statistical concepts**
  - **Run:** each observation made in the experiment
  - **Experimental error:** fluctuation in the results
  - Response variable is **random** due to the error
  - A random variable may be **discrete** or **continuous**

Simple comparative experiments

- **Review of basic statistical concepts**
  - **Probability distributions**: the probability structure of the random variable





  - **Mean, expected value, variance**

$$\mu = E(y) = \begin{cases} \int_{-\infty}^{\infty} y f(y)\, dy & y \text{ continuous} \\ \sum_{\text{all } y} y p(y) & y \text{ discrete} \end{cases} \qquad \sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\, dy & y \text{ continuous} \\ \sum_{\text{all } y} (y-\mu)^2 p(y) & y \text{ discrete} \end{cases}$$

Experimental Design and Hypothesis Testing

Simple comparative experiments

- **Sampling and sampling distributions**

The goal is to get conclusions about a population using a sample from that population

- **Random samples:** randomly taken from the population
- **Sample mean and variance:**

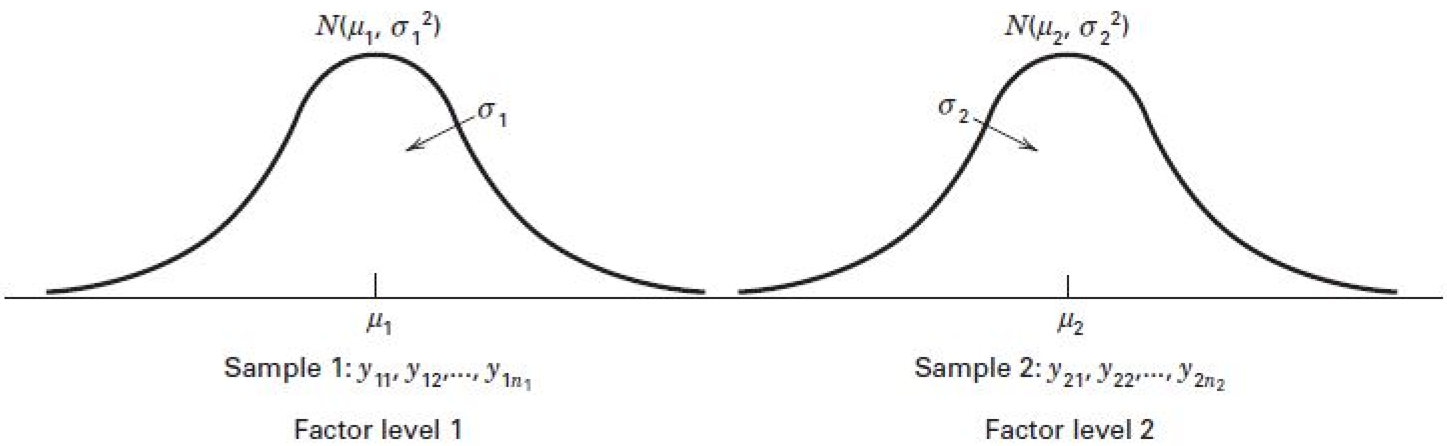
$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}$$

Experimental Design and Hypothesis Testing

Simple comparative experiments

- **Statistical Hypothesis**

A statement either about a parameter of a probability distribution or the parameter of the model (Portland cement example)



Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$ *(two-sided)* $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$.

Simple comparative experiments

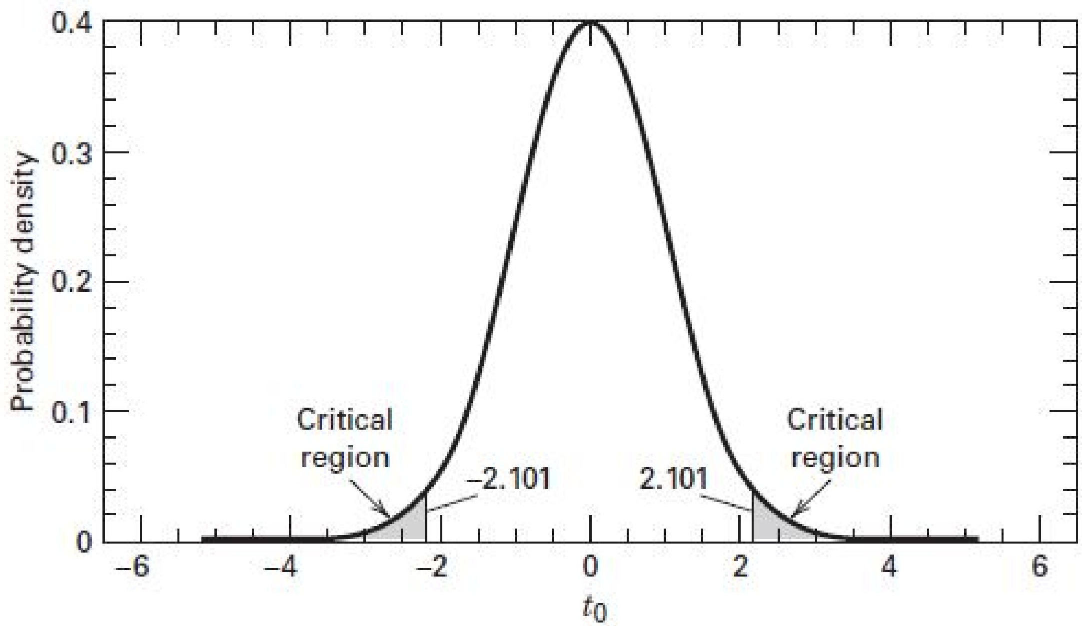
- **Testing a hypothesis**
  - Take a random sample
  - Compute an statistic test
  - Specify the critical/rejection region for the statistic test
  - Reject or not the null hypothesis based on the computed value of the statistic test
- **The two-sample t-test**

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Reject H_0 $|t_0| > t_{\alpha/2, n_1+n_2-2}$ Upper α/2 percentage point of the t-distribution with n_1+n_2-2 degree of freedom

Simple comparative experiments

- **The two-sample t-test (Portland cement example)**



t-distribution with 18 degree of freedom with a critical region

$$\pm t_{0.0025,18}$$

$$\boldsymbol{t_{0=-2.20} < -t_{0.0025,18} = -2.101}$$

Reject H_0

Simple comparative experiments

- **Confidence intervals**
  - Hypothesis testing is a useful technique, but sometimes is not enough.
  - An interval within which the value of the parameter(s) would be expected to lie is preferable. These interval statements are referred to as **confidence intervals.**

$$P(L \leq \theta \leq U) = 1 - \alpha$$

$$L \leq \theta \leq U \rightarrow$$

$$100(1-\alpha)\% \;\; confidence\ interval\ for\ \theta$$

  - **Portland cement example:** build a $100(1-\alpha)\%$ confidence interval for the difference in means of two treatments $(\mu_1 - \mu_2)$

Experimental Design and Hypothesis Testing

Simple comparative experiments

- **Confidence intervals (Portland cement example)**

$$\frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow t_{n_1+n_2-2}$$

$$P\left(-t_{\alpha/2,n_1+n_2-2} \le \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\alpha/2,n_1+n_2-2}\right) = 1 - \alpha$$

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2,n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2$$

$$\le \bar{y}_1 - \bar{y}_2 + t_{\alpha/2,n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Simple comparative experiments

- **Confidence intervals (Portland cement example)**

$$16.76 - 17.04 - (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}} \leq \mu_1 - \mu_2$$

$$\leq 16.76 - 17.04 + (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}}$$

$$-0.28 - 0.27 \leq \mu_1 - \mu_2 \leq -0.28 + 0.27$$

$$-0.55 \leq \mu_1 - \mu_2 \leq -0.01$$

- The **95% confidence interval** estimate of the difference in means of the bond strength varies from **-0.55 to -0.01**
- Since $\mu_1 - \mu_2 = 0$ does not fall into this interval, the data do not support the hypothesis that $\mu_1 = \mu_2$ with 5% level of significance!

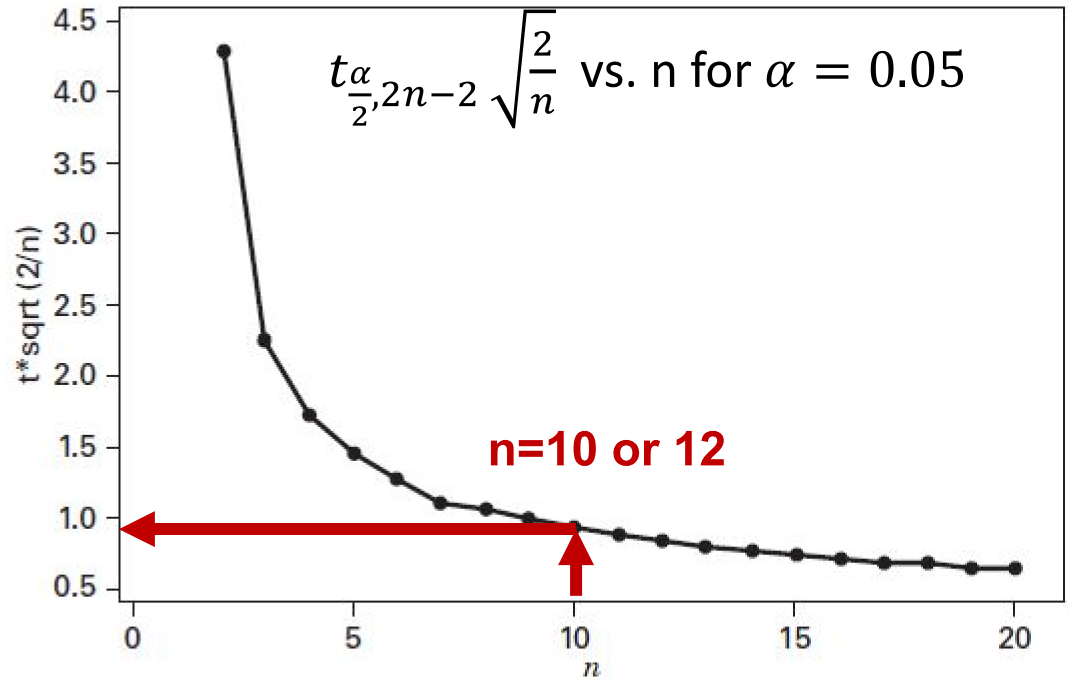
Simple comparative experiments

- **Choice of sample size**
  - Consider the impact of sample size on the estimate of the difference in the two means.

$$t_{\alpha/2,\, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

n1=n2

$$t_{\alpha/2,\, 2n-2}\, S_p \sqrt{\frac{2}{n}}$$



Error analysis of experimental results

- In most experiments, we need to compare 2 sets of results versus each other or versus a baseline which could be from an older experiment or a theoretical analysis.
- Depending on the results and their natural properties different approaches may be taken.
- We need to think about this ahead of time and plan for it.

Experimental Design and Hypothesis Testing

Error analysis of experimental results

- **Time independent results,**
  - The **amplitudes** of two datasets and their difference
  - the **absolute error value** and its **average** over the sample; or,
  - **Root mean square (RMS)** of the errors over the sample results.

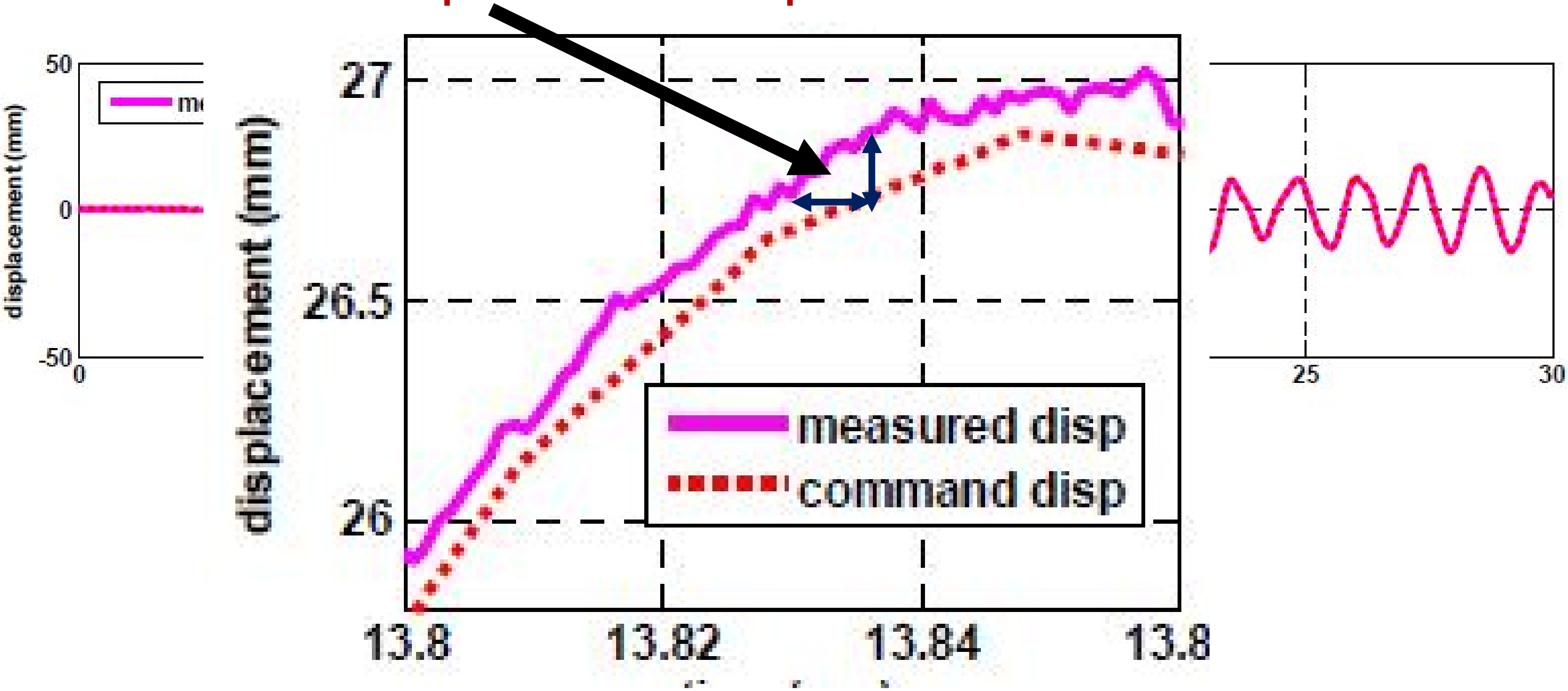
- **Time dependent results,**
  - Both **amplitude** and **phase** (lag or lead) **difference** in the results
  - The problem is phase (time) and amplitude errors are always coupled and it is a difficult task to decouple them.
  - Think about this issue ahead of time and come up with a method to tackle that

Experimental Design and Hypothesis Testing

Error analysis of experimental results

- **Time dependent results**

Phase and amplitude errors coupled

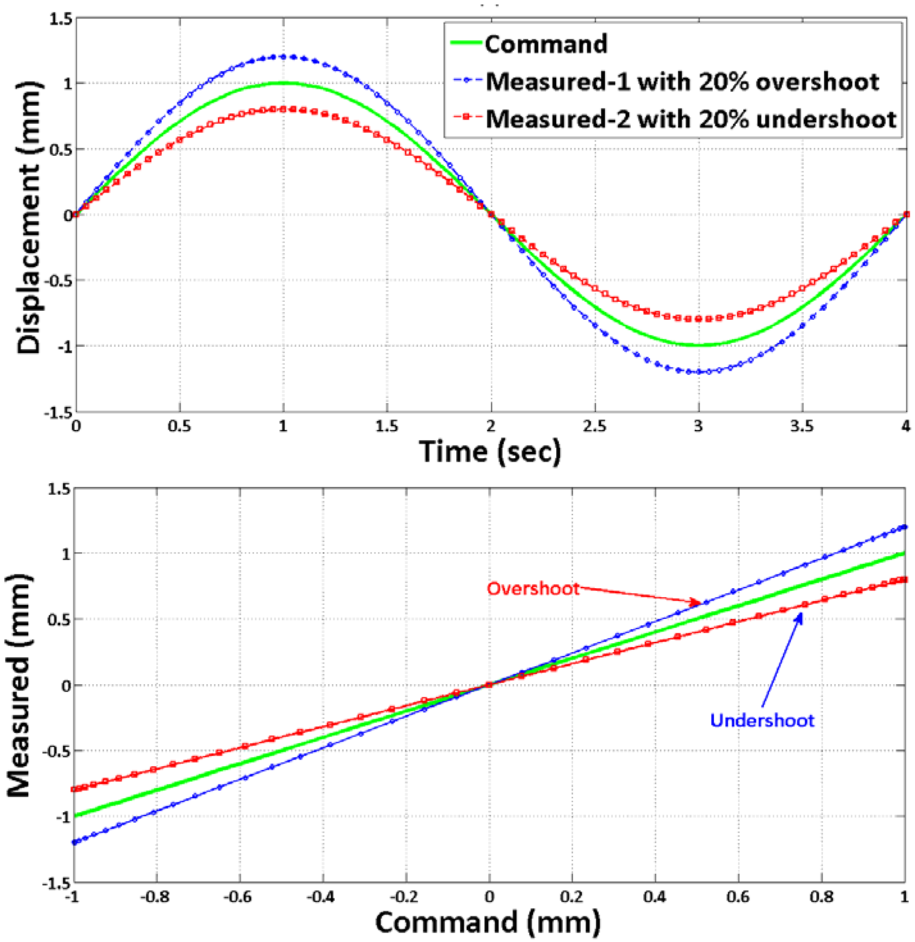


Experimental Design and Hypothesis Testing

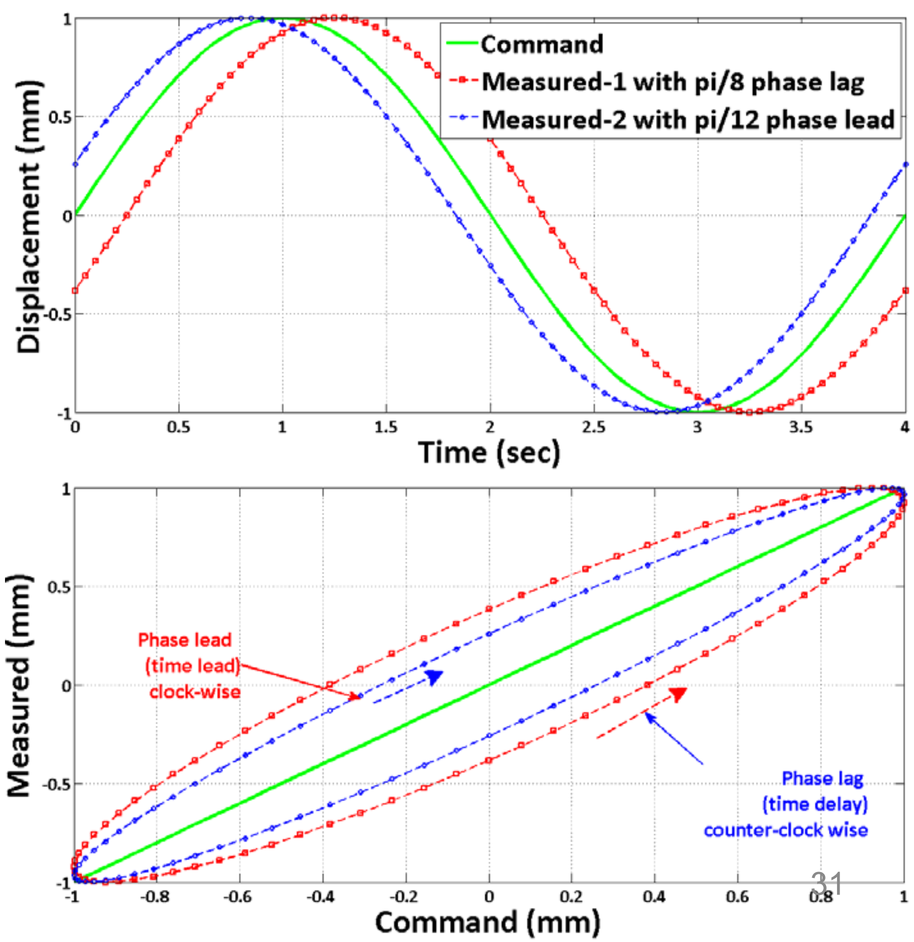
Error analysis of experimental results

- **Qualitative** methods of comparison **(synchronization subspace plots)**

Amplitude error



Phase (time) error



Experimental Design and Hypothesis Testing

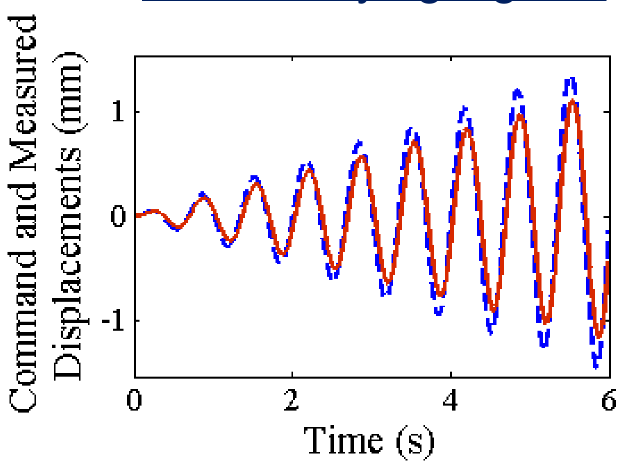
Error analysis of experimental results

- **Quantitative** methods of comparison

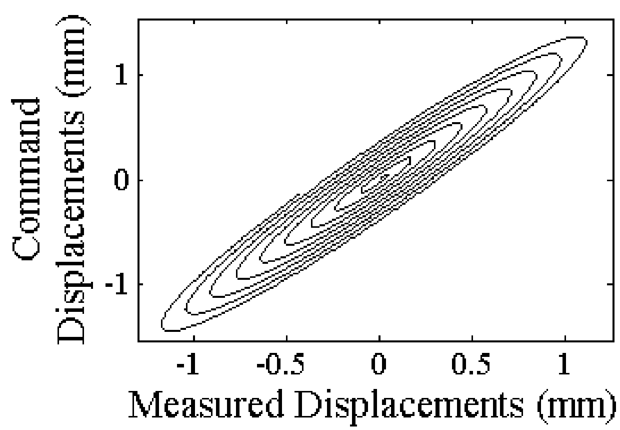
Error indicators for signals with time-varying amplitudes

- o **Tracking error (TI)**

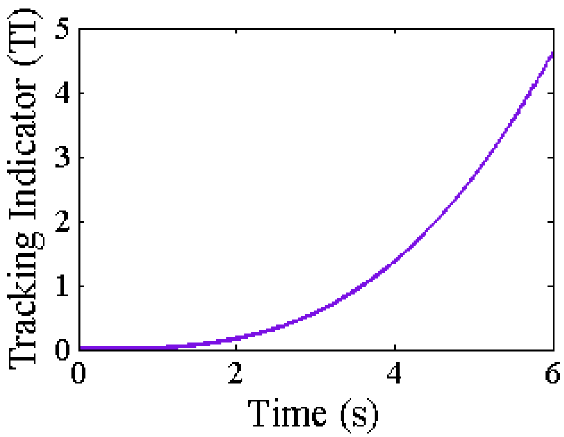
2 time-varying signals



SSP



Tracking Error (TI)



References

- Montgomery, D., Design and Analysis of Experiments, 8th edition, *John Wiley & Sons, Inc.,* 2012, ISBN 978-1118-14692-7
- Prof. Bryan Tolson class notes
- Prof. Liping Fu class notes
- Ashasi-Sorkhabi, A. "Implementation, Verification and Application of Real-time Hybrid Simulation." *PhD thesis*, University of Toronto, 2015

Questions ?