ESE

9. Data analysis and interpretation P2.2

09.2.1 Hypotheses Tests

Introduction

- Purpose of this presentation
- Background and definitions
- The “Great Table”

Presentation Purposes (1/3)

Let the following data come from observations:

Development time	
Method 1	Method 2
5	17

Development time	
Method 1	Method 2
5	17
12	30
50	100

Might a conclusion be drawn?

Presentation Purposes (2/3)

Development Time	
Method A	Method B
23	22
23	44
45	55
43	63
56	21
22	34
21	67
33	76
19	63
21	71
10	51
11	49
22	23
23	45
34	43
21	56
24	10

<

>

What should we conclude

?

Presentation Purposes (3/3)

In practice:

- How to run the **right** statistical test.
- Which **conclusion** can be drawn.

The Concept of Test (1/2)

- Very frequently a test aims to **reject** the hypothesis that two (or more) treatments had the **same** output.

The Concept of Test (2/2)

- Q: How can we judge two (or more) sets of data as equivalent or not?
  - Q1: In order to consider two sets of data as not equivalent, **how much** they have to be different?
  - Q2: **What** do we want to compare?
    - Means? Medians? Variances? ...

How much they have to be different?

Recalling the concepts of H_0 and H_1

- **H_0**:: the Null Hypothesis is used to refer to the state in which the two distributions are **not** significantly **different**.

- **H_1**:: the Alternative Hypothesis is used to refer to the state in which the two distributions are **significantly different**.

Errors in evaluating the Null Hypothesis

- **Error of type I**:: P(type-I-error)=

  P(reject H_0 | H_0 is true)

- **Error of type II**:: P(type-II-error)=

  P(NOT reject H_0 | H_0 is NOT true), i.e.,

  P(NOT reject H_0 | H_0 is false)

Error Notations α and β

- α :: P(Type I Error); **Level of significance** or **Size** of a test (result):

  The probability of a statistical test of incorrectly rejecting the (true) null hypothesis.

  The maximum probability with which we are prepared to run the risk of making a type I error.

  - 1–α: **Level of confidence** we have on rejection.
- β :: P(Type II Error). The probability of a statistical test of incorrectly accepting the (false) null hypothesis.
  - 1–β:: **Power of the statistical test**: the probability of a statistical test of correctly rejecting the (false) null hypothesis = 1-P(Type II Error).

The Concept of Test

For instance, we want an error probability α to be not greater than a given value, e.g.:

$$\alpha <= 0,10$$

The Concept of Test

Warning !!!

Which is the minimum level of α required to judge a distribution as different in respect to another?

It depends on our level of knowledge about the domain; e.g.:

$$\alpha <= 0.05$$

Normal vs. Not Normal Distribution

- We cannot reject that a distribution can be **considered** as normal if there was **no one test** able up to now to reject the null hypothesis of similarity in respect to a normal distribution.

Parametric vs. Not Parametric Tests

- Parametric tests can be used only when we cannot reject that each distribution object of the study is normal.

- Non parametric tests **can be used** instead of parametric tests but they are **less powerful**!

Overview of Tests (1/2)
One factor two treatments

- **t-test**: Often used parametric test to compare two samples *means*, i.e., the design is one factor with two treatments.
- **Mann-Whitney**: A non-parametric alternative to t-test.
- **F-test**: A parametric test that can be used to compare two sample distributions, in particular their *variance*.
- **Paired-t-test**: A t-test for paired comparison design.
- **Wilcoxon**: A non-parametric alternative to Paired t-test. It is based on *ranks* of the samples.
- **Sign-test**: A simple alternative to the Wilcoxon test. It is based on the *sign of differences* between samples.

Overview of Tests (2/2)

More than two levels of a factor.

- **ANOVA**: A family of parametric tests that can be used for designs with more than two levels of a factor. ANOVA test can, for example, be used in the following designs: One factor with more than two treatments, one factor with blocking variables, factorial design, and nested design.
- **Kruskal-Wallis** (or **One-way ANOVA on ranks**): A non parametric-test alternative to ANOVA in the case of one factor with more than two treatments.
- **Chi-2**: A family of non-parametric tests that can be used when data are in the form of *frequencies*.

The Great Table

Type of experiment	Parametric	Non Parametric
[redacted]		
Two Treatments, randomized	T-test	Mann-Whitney
	F-test	Chi Square
Two treatments, paired	Paired T-test	Wilcoxon
		Sign test
Three or more tratments	ANOVA	Kruskal-Wallis
		Chi Square

Paired Design

It is the instance of the **Completely Randomized Complete Block Design** for the case of “One factor two treatments”.

1. **Paired Design**: each subject uses **both treatments on each object**.
   - Be sure to balance the order and that the learning effect is not destructive.