Data Set Reduction and Hypothesis Testing

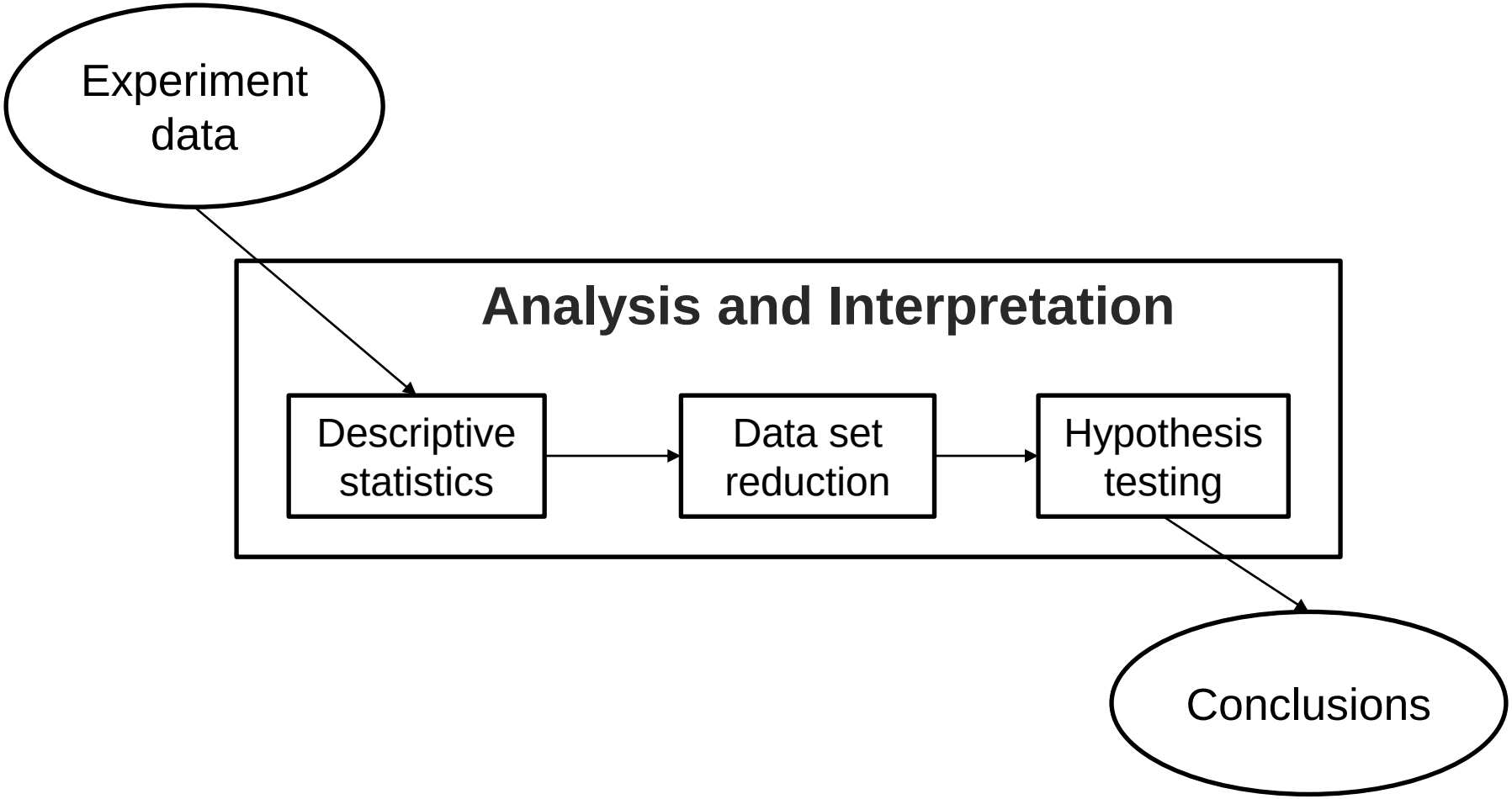
Eduardo Figueiredo

http://www.dcc.ufmg.br/~figueiredo

Analysis and Operation

- After collecting data in the operation phase, we want to draw conclusions
  - The analysis and operation phase aims to interpret the collected experimental data
- This phase has three main steps
  - Descriptive statistics
  - **Data set reduction**
  - **Hypothesis testing**

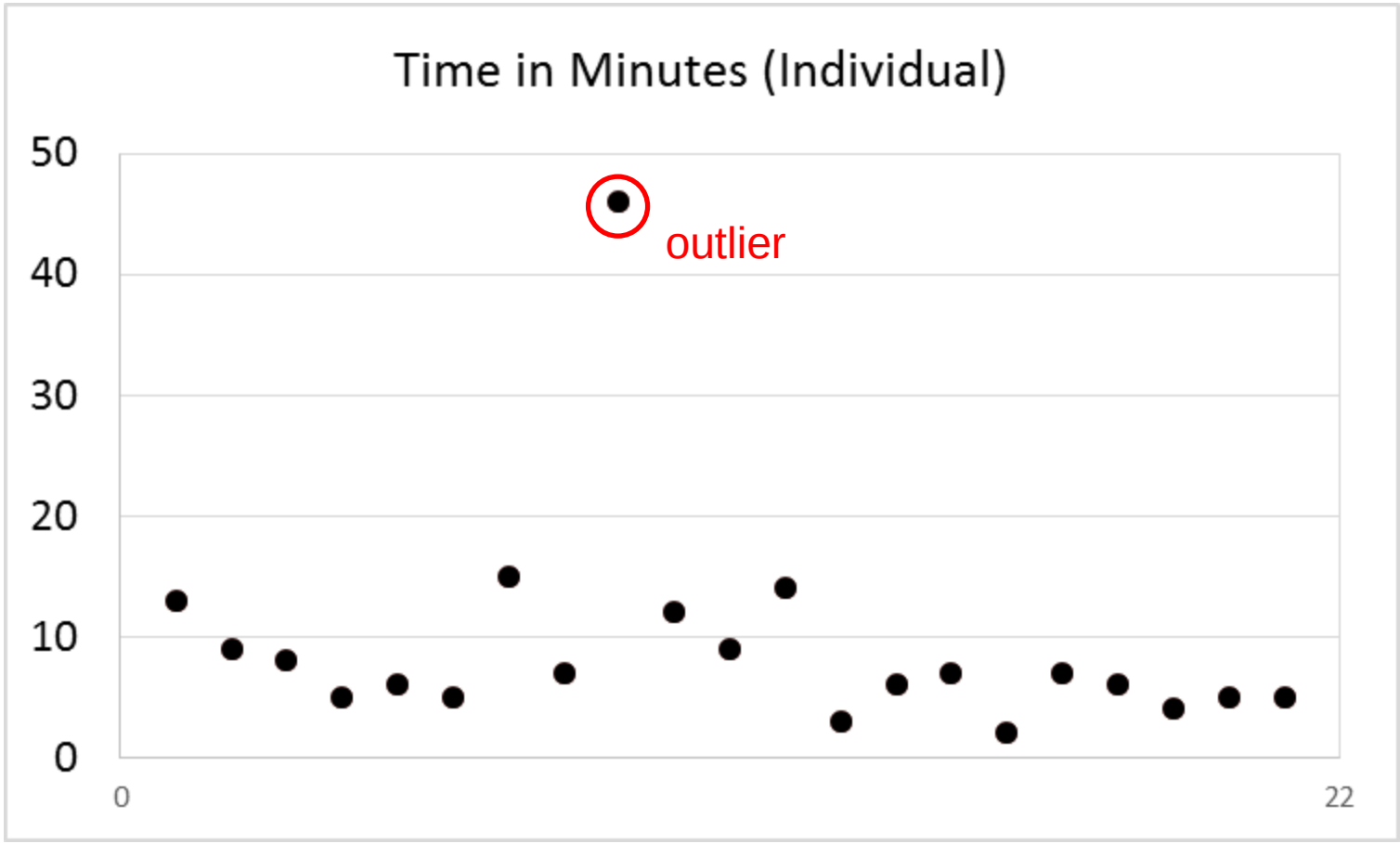
Analysis and Operation Overview



Data Set Reduction

- Analysis and interpretation depend on the quality of the data
  - If data include measurement errors, conclusions may be incorrect
- Unexpected outliers can often be removed from the data set
  - You should remove only a strange or rare value that was never expected to happen

Example of Outlier to Remove



Hypothesis Testing

- Its goal is to see if it is possible to reject a null hypothesis (H_0)
  - The experimenter wants to reject H_0 with a given significance
- Example
  - H_0: Work experience does not impact on productivity
  - H_1: Work experience impacts on productivity

Example of Experiment

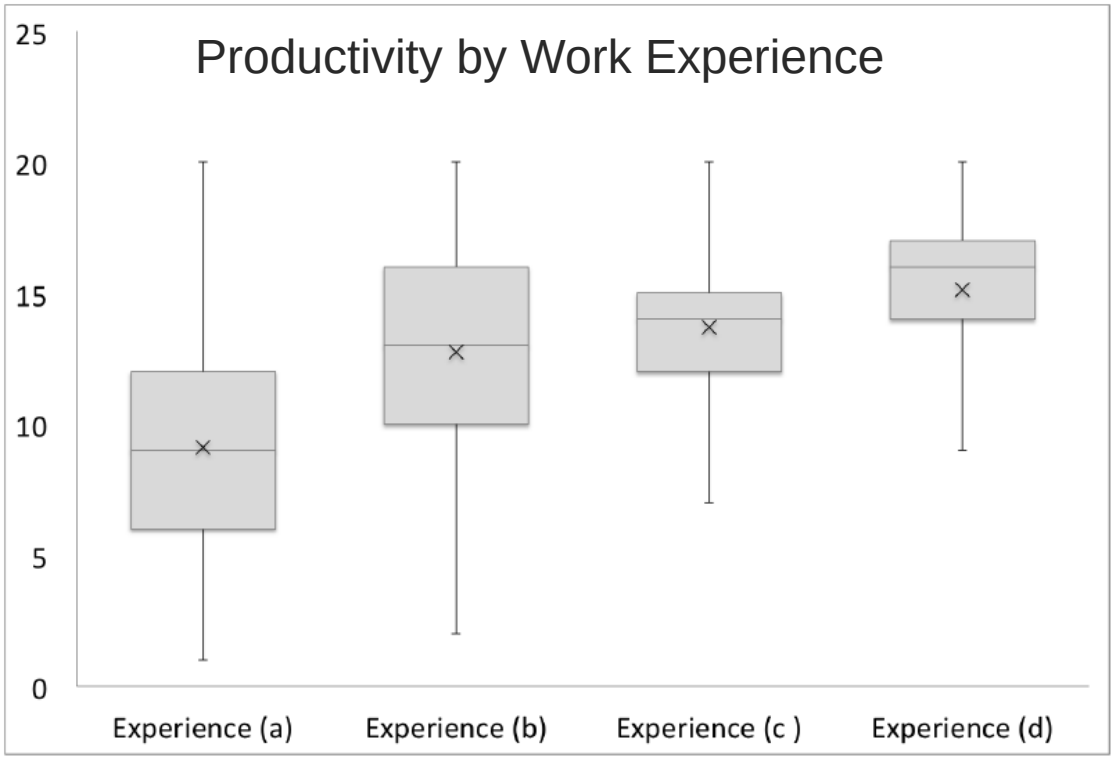


Figure	n	Mean	Std. deviation	Min	Q1	Q2	Q3	Q4
(a)	6071	9.10	3.71	1	6	9	12	20
(b)	704	12.74	3.89	2	10	13	16	20
(c)	551	13.70	2.15	7	12	14	15	20
(d)	158	15.13	2.13	9	14	16	17	20

Rejecting H_0

- If the null hypothesis is not rejected, nothing can be said about the outcome
  - If the null hypothesis is rejected, the opposite is true with a given significance α
- When a test is executed, it is often possible to calculate the lowest significance (p-value)
  - If p-value is lower than α, it is possible to reject the null hypothesis

Power of Test

- Power of test is the probability of rejecting H_0 when it is false
  - Type-I-error: rejecting a true H_0
  - Type-II-error: do not rejecting a false H_0
- Factors affects the power of test
  - The test can be less effective
  - The sample size (small means less power)

Parametric and Non-Parametric

- Tests can be classified as parametric and non-parametric tests
- Parametric tests
  - They assume normal distribution
  - One test of normality is Chi-2
  - They require at least interval scale
- Non-parametric tests
  - They are more general and do not assume distribution

Factors for the Test Choice

- Applicability
  - Distribution and scale type have to be realistic
  - Restrict the choice of parametric tests
- Power
  - The power of parametric tests are higher
  - Parametric tests require fewer data points (suitable for small experiments)

Overview of Tests

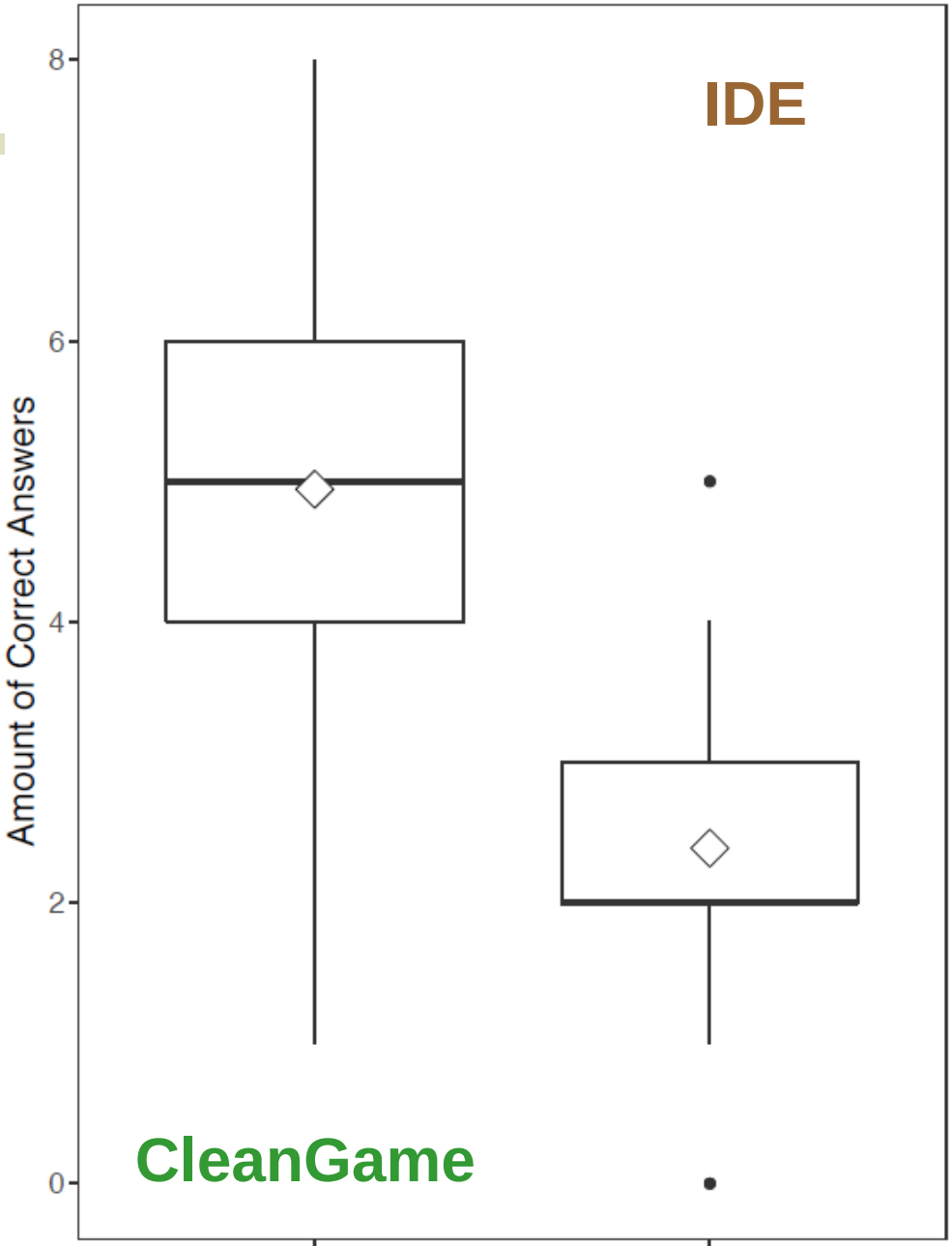
- Common parametric tests
  - t-test
  - F-test
  - Paired t-test
  - ANOVA
- Common non-parametric tests
  - Mann-Whitney
  - Wilcoxon
  - Sign test
  - Kruskal-Wallis
  - Chi-2

Tests for Different Designs

Design	Parametric	Non-parametric
One factor, one treatment		Chi-2, Binomial
One factor, two treatments (random)	t-test, F-test	Mann-Whitney, Chi-2
One factor, two treatments (paired)	Paired t-test	Wilcoxon, Sign test
One factor, more than two treatments	ANOVA	Kruskal-Wallis, Chi-2
More than one factor	ANOVA	

Example SBES 2019

- **Paired Wilcoxon's** rank-sum test
  - non-parametric
- Subjects perform significantly better using CleanGame
  - p-value = 0.003



Bibliography

- C. Wohlin et al. **Experimentation in Software Engineering**, Springer. 2012.
  - Chapter 10 – Analysis and Interpretation (Section 10.2 Data Set Reduction and Section 10.3 Hypothesis Testing)