Experiment Design and Statistical Data Analysis

Dr. Pradipta Biswas

Senior Research Associate, Department of Engineering Fellow, Wolfson College University of Cambridge, UK

E-mail: pb400@cam.ac.uk
Web: http://www-edc.eng.cam.ac.uk/~pb400

Structure for today

Introduction
- Experiment Design
- Data Preparation
- Test Selection

Discussion
- Your projects
- Data analysis

Details on tests
- T-Test
- ANOVA
Histogram

Central Tendency

- mode
- median
- mean

$\sigma = 0.25$

$\sigma = 1$
How to test

• Sampling: Select a set of participants
• Method: Design a study to collect data
• Material: Get instruments
• Procedure: Collect data from participants
• Result: Analyze data

Sampling

• Size
  – No straight forward answer!!
  – Can be estimated statistically
  – Bigger the better
    • More representative of population
  – Often limited by availability

• Quality
  – Random sampling
  – Group based sampling
  – Purpose based sampling
Variables

• Variables are things that change

• The **independent variable** is the variable that is purposely changed. It is the manipulated variable.

• The **dependent variable** changes in response to the independent variable. It is the responding variable.

Constant Variables

• Factors that are kept the same and not allowed to change.

• It is important to control all but one variable at a time to be able to interpret data
Hypothesis

• Your best thinking about how the change you make might affect another factor.

• Tentative or trial solution to the question.

• An if ............ then ............ statement.

• Should be expressed in measurable terms

Experiment

• Random assignment
  – Factorial design: More than one IV
  – Parametric design: IV has more than two levels

• Matched pair

• Repeated measure
Data Screening

• Skewing
  – In opposite direction

• Unequal Variance
  – Equal number of samples: $\frac{\sigma^2_{\text{max}}}{\sigma^2_{\text{min}}} < 4$
  – Unequal number of samples: $\frac{\sigma^2_{\text{max}}}{\sigma^2_{\text{min}}} < 2$

• Random Error

• Missing Values

• Data Transformation

Normality Check

We should check for normality using:
• assumptions about population
• histograms for each group
• normal quantile plot for each group

With such small data sets, there really isn’t a really good way to check normality from data, but we make the common assumption that physical measurements of people tend to be normally distributed.
Test selection

- Data normally distributed
  - Parametric / Non-parametric

- Relationship between two columns of data
  - Correlation (Pearson’s \( r \) / Spearman’s \( \rho \))

- Comparing means between two columns of data
  - T-test / U-Test / Wilcoxon signed rank test

- More than two columns
  - ANOVA / Kruskal-Wallis H test

Scatter plot & Correlation

![Scatter Plot](image)
Correlation - outliers

$r = 0.66$

$r = -0.27$

Error plot

Relative Error in Prediction

0 2 4 6 8 10 12 14 16 18
0 10 20 30 40 50 60 70 80 90 100
<-120 -100 -80 -60 -40 -20 0 20 40 60 80 100 120
% Error
% Data
Important terms

• Degrees of freedom (df)

• One tail and two tail tests
  – Better/Worse or just different

• Type I (α) and Type II (β) error

• Sphericity assumption (for ANOVA)

Comparing means – t-test
The basic ANOVA situation

Two variables: 1 Categorical, 1 Quantitative

Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?

If categorical variable has only 2 values:
• 2-sample t-test

ANOVA allows for 3 or more groups

An example ANOVA situation

Subjects: 25 patients with blisters
Treatments: Treatment A, Treatment B, Placebo
Measurement: # of days until blisters heal

Data [and means]:
• A: 5,6,6,7,7,8,9,10 [7.25]
• B: 7,7,8,9,9,10,10,11 [8.875]
• P: 7,9,9,10,10,10,11,12,13 [10.11]

Are these differences significant?
Informal Investigation

Graphical investigation:
- side-by-side box plots
- multiple histograms

Whether the differences between the groups are significant depends on
- the difference in the means
- the standard deviations of each group
- the sample sizes

ANOVA determines P-value from the F statistic

Side by Side Boxplots
What does ANOVA do?

At its simplest (there are extensions) ANOVA tests the following hypotheses:

H\(_0\): The means of all the groups are equal.

H\(_a\): Not all the means are equal
  • doesn’t say how or which ones differ.
  • Can follow up with “multiple comparisons”

Note: we usually refer to the sub-populations as “groups” when doing ANOVA.

Assumptions of ANOVA

• each group is approximately normal
  ① check this by looking at histograms and/or normal quantile plots, or use assumptions
  ② can handle some nonnormality, but not severe outliers

• standard deviations of each group are approximately equal
  ① rule of thumb: ratio of largest to smallest sample st. dev. must be less than 2:1
Standard Deviation Check

<table>
<thead>
<tr>
<th>Variable</th>
<th>treatment</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>days</td>
<td>A</td>
<td>8</td>
<td>7.250</td>
<td>7.000</td>
<td>1.669</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>8</td>
<td>8.875</td>
<td>9.000</td>
<td>1.458</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>9</td>
<td>10.111</td>
<td>10.000</td>
<td>1.764</td>
</tr>
</tbody>
</table>

Compare largest and smallest standard deviations:
- largest: 1.764
- smallest: 1.458
- $1.458 \times 2 = 2.916 > 1.764$

Note: variance ratio of 4:1 is equivalent.

How ANOVA works (outline)

ANOVA measures two sources of variation in the data and compares their relative sizes

- variation BETWEEN groups
  - for each data value look at the difference between its group mean and the overall mean
    \[
    (\bar{x}_i - \bar{x})^2
    \]

- variation WITHIN groups
  - for each data value we look at the difference between that value and the mean of its group
    \[
    (x_{ij} - \bar{x}_i)^2
    \]
Result

\[ \text{Test Statistics} = \frac{\text{Variance explained by model}}{\text{Variance not explained by model}} = \frac{\text{Effect}}{\text{Error}} \]

Significant \(\rightarrow\) **The probability that the model is explaining variance by chance < 0.05**

F- statistics

The ANOVA F-statistic is a ratio of the Between Group Variation divided by the Within Group Variation:

\[ F = \frac{\text{Between}}{\text{Within}} = \frac{\text{MSG}}{\text{MSE}} \]

A large F is evidence against \(H_0\), since it indicates that there is more difference between groups than within groups.
ANOVA Output

Analysis of Variance for days

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.006</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MS = SS /df

F = MSG /MSE

Effect size and Power

- **Effect size**
  - Percent of variance explained
  - Standardized measure of magnitude of effect
  - Cohen’s $d$, correlation coefficient, $\eta^2$

- **Power**
  - Power of a test to detect significant effect
  - $(1 – \text{Type II error})$
    - Type II error ($\beta$) $\rightarrow$ probability of not detecting an effect
    - Can be used to estimate sample size
Other tests

Other important tests won’t be discussed in detail but relevant to HCI trials

– Non Gaussian distribution → Non parametric tests

– Comparing ranks → Sign test

– ANCOVA

– MANOVA and so on

Reporting

• Title
• Abstract
• Introduction
• Method
  – Participants
  – Materials
  – Design
  – Procedure
• Results
• Discussion
• References
• Appendix (Optional)
Take away points

• Introduction to the process of designing a study or experiment and conducting user trial

• Basic data screening and analysis techniques

• Basic statistical methods and terms associated with conducting controlled experiment

• Reporting a study following standard format