S 08

Data Analytics for Data Scientists

Design of Experiments (DoE)

Suggested solutions for 08: A/B Testing

2025

Prof. Dr. Jürg Schwarz
MSc Adrian Bieri

Suggested solution 01

Properties

What are the characteristics of "classical" A/B testing procedures and of the dynamic-algorithmic procedure?

Please use one example for the "classical" A/B testing and one for a dynamic-algorithmic procedure and describe the advantages and disadvantages for each.

Suggested answers to the questions

Example of A/B testing	**Example of dynamic-algorithmic procedure**
Description Three variants of a website are tested with regard to their effect on the conversion rate. Three independent samples are randomly selected from website visitors during the measurement period and assigned to the three variants. The conversion rate is measured in the three variants.	Description Three variants of a website are tested with regard to their effect on the conversion rate. A multi-armed bandit algorithm is used for each run to dynamically select the variants with the highest conversion rate. These steps are repeated until a stop rule is applied.
Advantages The three variants can be compared by means of ANOVA, making it possible to show which of the variants has the highest, second highest, and third highest conversion rate (post-hoc test).	Advantages The test delivers a result quickly.
Disadvantages Depending on the prerequisites (number of variants, statistical analysis (ANOVA, etc.), level of alpha error), many observations are required according to power analysis.	Disadvantages There is "only" one result that has been produced by the dynamic process. This means that no models can be estimated, which for example test the dependence of the conversion rate on other factors.

Suggested solution 02

An example

A multivariate test for a website was carried out.
The test measured the average click rate (number of clicks on advertising banners) on a website.

How do the design variants (factor IV1) affect the click rate (DV)?

DV	=	click rate [s]	→ output with metric scaling
IV1	=	design variants [1,2,3]	→ factor with 3 levels

Use a univariate ANOVA to answer the question.

Describe the results – insert also the R-code, the R-output and if necessary, R-plots.

Data set: clickrate.xlsx → ILIAS → "Data Resources & R Files"

Suggested answers to the questions

```
library(readxl)
library(effectsize)

clickrate <- read_excel("clickrate.xlsx")

fit  <-  aov(DV  ~  factor(IV1),  data  =  clickrate)
summary(fit)

pairwise.t.test(clickrate$DV, clickrate$IV1, p.adj = "bonf")

aggregate(clickrate$DV, list(clickrate$IV1), mean)

eta_squared(fit)$Eta2
```

```
              Df Sum Sq Mean Sq F value Pr(>F)    
factor(IV1)    2  13256    6628   77.93 <2e-16 ***
Residuals   2301 195708      85                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
   Pairwise comparisons using t tests with pooled SD 

data:  clickrate$AV and clickrate$UV1 

  1       2      
2 2.2e-07 -      
3 < 2e-16 7.1e-12

P value adjustment method: bonferroni
```

```
  Group.1        x
1       1 36.23021
2       2 38.77109
3       3 42.08854
```

```
[1] 0.06343711
```

There is a main effect of IV1 (1, 2, 3) on DV, $F(2, 2301) = 77.93$, $p = .000$.

Design variants (1, 2, 3) have a significant effect on the click rate.
The mean of the click rate in design variant 1 is 36.2 seconds, in variant 2 38.8 seconds, and in variant 3 42.1 seconds.
All means differ in pairs (Post-hoc Bonferroni).

Effect size f for the *One-way analysis of variance* (Cohen 1992), calculated from η_p^2

$$f = \sqrt{\frac{\eta_p^2}{1-\eta_p^2}} = \sqrt{\frac{0.063}{1-0.063}} = \sqrt{\frac{0.063}{0.937}} = 0.26$$

	Small	Medium	Large
Effect size *f*	0.10	0.25	0.40

The effect size is $f = 0.26$ (Cohen 1992)