

Comparing groups

Marketing Research and Analytics



WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Andreas MILD

Learning objectives

- Comparison between groups
- Tables
- Visualizations
- Tests

Questions & Dataset

- Marketing analytics often investigates differences between groups
 - Do men or women subscribe to a service at a higher rate?
 - Which demographic segment can best afford a product?
 - Does region A perform better than region B?
 - Did same-store sales increase after a promotion?
- Answers help to
 - understand the market
 - target customers more effectively
 - evaluate the outcome of marketing activities such as promotions
- Data from a consumer segmentation project
- N=300
- Variables:
 - Age, Gender, Income, # children
 - own or rent a home
 - subscribe yes/no
 - Customer segment: Suburb mix/Urban hip/Travelers/Moving up

21 The Cosmopolitans



Upscale Younger Family Mix

Age 25-44 Homeowners Moderate IPA Below Average Tech

Educated and upscale, The Cosmopolitans are urbane couples in America's fast-growing cities. Concentrated in major metro areas, a vibrant social scene surrounds their older homes. These residents tend to live an eco-friendly lifestyle and enjoy eating at organic/health food establishments.



Owns a BMW • Eats at organic/health food restaurants • Shops online at Crate & Barrel • Attends PAC-12 sports events • Visits Europe • Watches the Video Music Awards • Listens to Album Adult Alternative

[Lifestage Group: Y1- Midlife Success \(Younger Years\)](#)

[Social Group: U1- Urban Uptown \(Urban\)](#)

Comparing groups: statistical tests

- Visual inspection and tables let us discover differences in groups
- „It looks different, but is it really different?“
- Inferential statistical procedures:
 - Testing frequencies: `chisq.test()`
 - Testing observed proportions: `binom.test()`
 - Testing group means: `t.test()`
 - Testing multiple group means: `anova()`

Confidence intervals and significance

- Definition of a 95% confidence interval:
 - It is the range of possible estimates that we would expect to see 95% of the time if we repeatedly estimate a statistic using *random samples* of the *same sample size* under the assumption that the *true value* in an (almost) *infinite population* is the same as our current estimate
- Confidence intervals do not tell us how confident we are in the answer
- Non significant results:
 - Evidence for the result is weak and should not be interpreted
 - More data might be collected
- Significant results:
 - Significance is no measure for the (practical) importance of the result