

Marketing & Data Analysis

Statistical Comparisons

Santiago Alonso-Díaz

Tecnológico de Monterrey
EGADE, Business School



Photo: Dalle2

How do we know our customers?

¿Qué prefiere azulito?

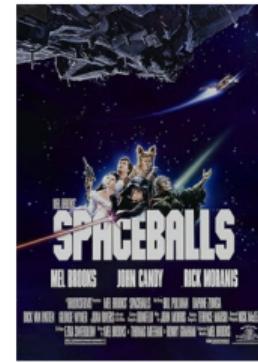
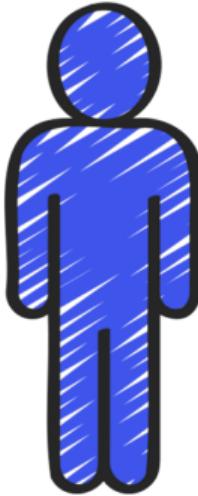


Figure: No data. Does azulito prefers comedies or thrillers?

How do we know our customers?



Figure: With data. Does azulito prefers comedies or thrillers?

How do we know our customers?

Dada la anterior data
¿Qué prefiere azulito?

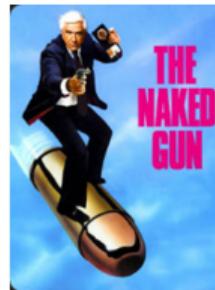
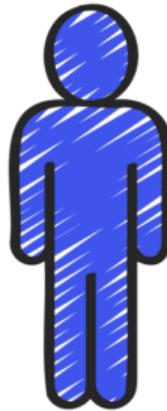


Figure: Given our prior data. What would azulito pick?

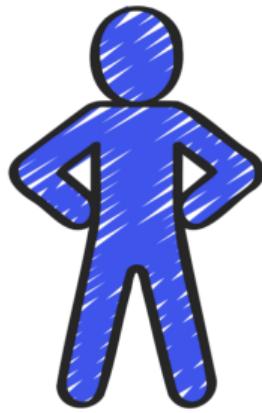
How do we know our customers?



Figure: Which scenario would increase your confidence of your previous inferences about azulito?

How do we know our customers?

Dada la data anterior
¿A quién prefiere conocer azulito?



Película favorita:
12 monkeys



Película favorita:
Loca academia de policía

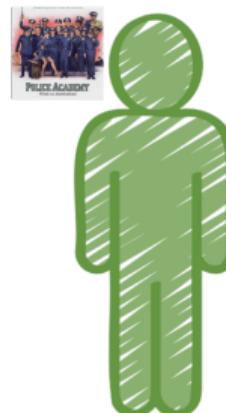


Figure: Can/should we make higher order inferences?

How do we know our customers?

What do we need to characterize azulito?

- Demographical data
- Choice data
- Preference data
- Neuro-cognitive-emotional data
- Theory (Mathematical, Statistical, Conceptual)

Let's see a real world example

How do we know a campaign worked in social media?

Which is better?

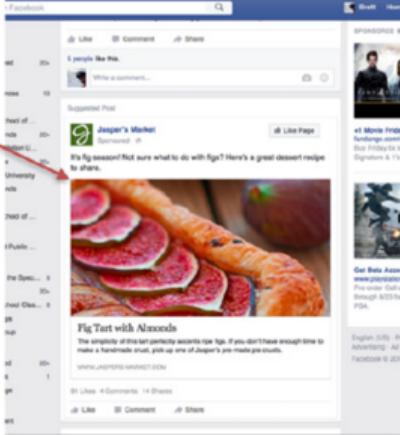
- Experimental data (RCT).
- Observational data (e.g. sales)
- Hearsay ("my social bubble told me so")
- Intuition ("I feel it works")
- Tradition ("it always works")

Experiment Approach

Ad Auction

1. 
2. 
3. 
4. 

Test

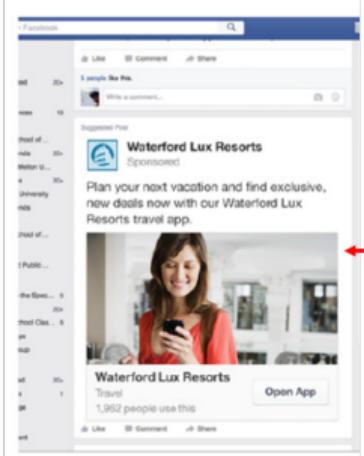


The image shows a Facebook news feed with four ads. The first ad is from 'J' (green logo). The second ad is for 'Fig Tart with Almonds' (pink logo), which is the test campaign. The third ad is from 'Waterford Lux Resorts' (blue logo). The fourth ad is from 'LUXE GENEVA' (purple logo). A red arrow points from the 'Fig Tart' ad to the text '2nd best ad'.

Ad Auction

1. 
2. 
3. 
4. 

Control



The image shows a Facebook news feed with four ads. The first ad is from 'J' (green logo). The second ad is for 'Waterford Lux Resorts' (blue logo), which is the control campaign. The third ad is from 'Fig Tart with Almonds' (pink logo). The fourth ad is from 'LUXE GENEVA' (purple logo). A red arrow points from the 'Waterford Lux Resorts' ad to the text '2nd best ad'.

Figure: Control: if the test campaign wins the auction, it is the 2nd best ad the user sees instead of the test campaign (Gordon et al., 2019). Is it a good randomization strategy?

Alonso-Díaz

Marketing & Data Analysis

9 / 17

Observational Approach

Usually a regression. For instance

$$\text{Checkout} = \beta_0 + \beta_1 \text{Exposure}_{add} + \theta \text{Demographics}$$

There are now techniques to pull causality from these type of regressions. Gordon et al., 2019 use those techniques to compare them with an RCT.

How do we know a campaign worked? Stats

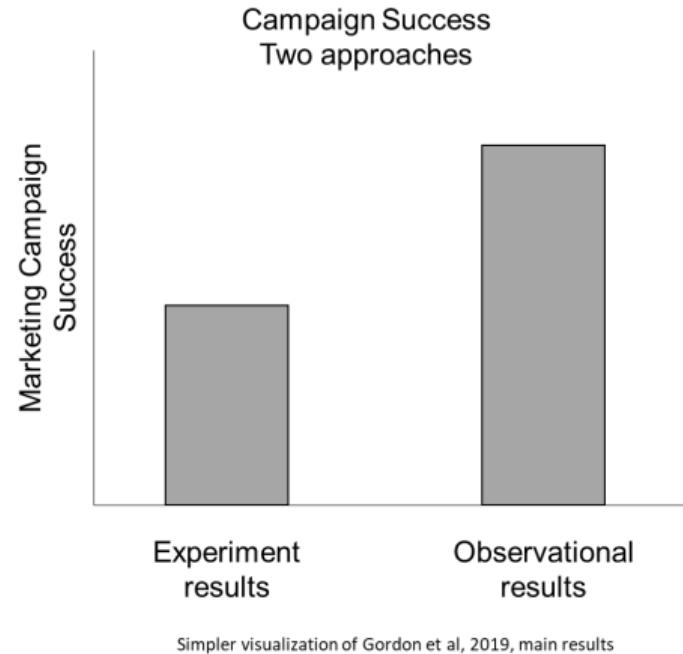


Figure: Experiment or observations? Both indicate success.

Table of contents

1 Overview

2 Testing Hypotheses

- Proportion comparisons
- Mean comparisons (t-tests, ANOVA)

3 References

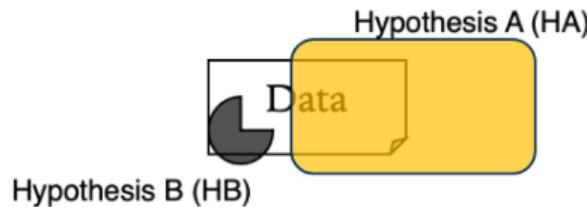
Testing Hypotheses

Testing Hypotheses

Data allow us to test hypotheses relevant to marketing:

- Is azulito a thriller movies fan?
- Does an ad campaign increases sales in social media?

Two approaches



Approach 1:
max. $p(\text{Data}|H)$

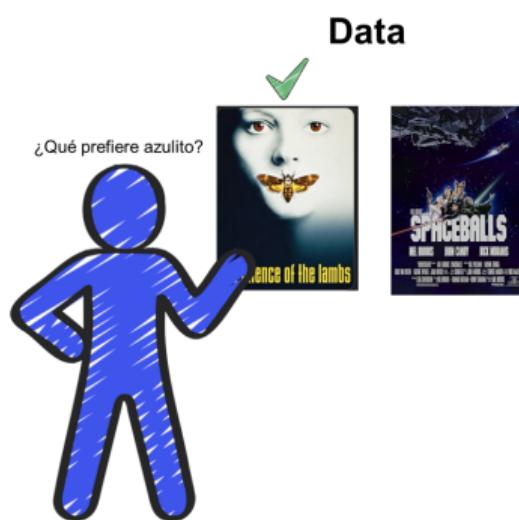
Given HB, the data covers most of HB. Pick HB

Approach 2:
max. $p(H|\text{Data})$

Given the data, HA covers more data. Pick HA

Figure: Maximize probability of data or hypotheses?

Example



Approach 1

$p(\text{Silence of the Lambs} | \text{Thriller Fan}) = ?$

$p(\text{Silence of the Lambs} | \text{Comedy Fan}) = ?$

Approach 2

$p(\text{Thriller Fan} | \text{Silence of the Lambs}) = ?$

$p(\text{Comedy Fan} | \text{Silence of the Lambs}) = ?$

Figure: Are these probabilities the same? Do they answer the same questions?

Two approaches

In this course we will focus on the first approach: Maximize the probability of the data given a hypothesis

But a fundamental question: can we always compute probabilities?? Yes but we make assumptions. Look for these distributions in Internet:

- Heights
- Age
- Income in Mexico

Let's go to Python

FMDA_4_Inference.ipynb

References



Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2), 193–225.