

# Razor market inspection

Group 8 - Khoa D. Nguyen

## 1. State the Question/Problem

- a. Hypothesis for Response and Turnover Rates across Promotional Channels:
  - Null Hypothesis (H0): There is no difference in response and conversion rates across promotional channels when receiving promotional coupons, specifically SMS and Email, among any age groups.
  - Alternative Hypothesis (H1): There is a difference in response and conversion rates across promotional channels when receiving promotional coupons, specifically SMS and Email, among any age groups.
- b. Hypothesis for Response and Turnover Rates Across Age Groups:
  - Null Hypothesis (H0): There is no difference in response and conversion rates across age groups (15-30, 31-45, 46-60, 60+) when receiving promotional coupons via both SMS and Email or via a single channel.
  - Alternative Hypothesis (H1): There is a difference in response and conversion rates across age groups (15-30, 31-45, 46-60, 60+) when receiving promotional coupons via both SMS and Email or via a single channel.
- c. Hypothesis for Experience Across Age Groups:
  - Null Hypothesis (H0): There is no difference in experience variables across age groups (15-30, 31-45, 46-60, 60+).
  - Alternative Hypothesis (H1): There is a difference in experience variables across age groups (15-30, 31-45, 46-60, 60+).

## 2. Define Hypothesis in Statistical Terms

- a. Hypothesis for Response and Turnover Rates across Promotional Channels:

<b>Null Hypothesis (H0)</b>	$\mu_{\text{SMS}}^i = \mu_{\text{Email}}^i$ for i in (15-30, 31-45, 46-60, 60+)
<b>Alternative Hypothesis (H1)</b>	$\mu_{\text{SMS}}^i \neq \mu_{\text{Email}}^i$

	for i in (15-30, 31-45, 46-60, 60+)
--	-------------------------------------

b. Hypothesis for Response and Turnover Rates Across Age Groups:

<b>Null Hypothesis (H0)</b>	$\mu_m = \mu_n$ for m, n in (15-30, 31-45, 46-60, 60+)
<b>Alternative Hypothesis (H1)</b>	$\mu_m \neq \mu_n$ for m, n in (15-30, 31-45, 46-60, 60+)

c. Hypothesis for Experience Across Age Groups:

<b>Null Hypothesis (H0)</b>	$\mu_m = \mu_n$ for m, n in (15-30, 31-45, 46-60, 60+)
<b>Alternative Hypothesis (H1)</b>	$\mu_m \neq \mu_n$ for m, n in (15-30, 31-45, 46-60, 60+)

### 3. Describe the Data Source

Data source description (metadata):

<b>Who</b>	Trusting Social SEA
<b>When</b>	2019-04-01 to 2019-05-01
<b>What</b> variables are included	'id', 'send_date', 'estimated_age', 'age_range', 'channel', 'coupon', 'clicked', 'last_step', 'nb_units', 'order_value', 'prev_price', 'replace_time', 'satisfaction_score'

<b>Why</b>	Identify the main customer groups and preferred method of promotion and communication for different customer groups
<b>How</b> it was collected	Data collected from an in-house ecommerce website's database.

#### Data view:

- Coupon interaction data columns (total 10 columns):
 

#	Column	Non-Null Count	Dtype
0	id	300000 non-null	int64
1	send_date	300000 non-null	object
2	estimated_age	300000 non-null	int64
3	age_range	300000 non-null	object
4	channel	300000 non-null	object
5	coupon	300000 non-null	int64
6	clicked	300000 non-null	int64
7	last_step	300000 non-null	object
8	nb_units	300000 non-null	int64
9	order_value	300000 non-null	int64
- Survey data columns (total 6 columns):
 

#	Column	Non-Null Count	Dtype
0	id	17766 non-null	int64
1	age_range	17766 non-null	object
2	last_step	17766 non-null	object
3	prev_price	17766 non-null	int64
4	replace_time	17766 non-null	int64
5	satisfaction_score	17766 non-null	int64

## 4. Define Sampling Criteria (if Applicable)

Sampling category: None, although given the set-up of the data collection process, only Southeast Asian male will be participating.

## 5. Specify Sample Size and Number of Variables

- a. Coupon interaction data:
  - Sample size: 300000 coupons sent / 31 days observed.
  - Number of variables: 6 variables
- b. Survey data:
  - Sample size: 17766 website interaction / 31 days observed
  - Number of variables: 6 coupon variables + 3 survey variables

## 6. Create a Data Dictionary

id	Format: Integer, representing each message
send_date	Format: data, date when sms/Email was sent
estimated_age	Format: Integer, ranging from 0 to 100
age_range	Format: string. Audience is divided into 4 age ranges
channel	Format: string, either SMS or Email
coupon	Format: float, the value of coupon expressed in each message, valid for up to 3 units for each order
clicked	Format: binary, either 0 (customer doesn't click on the link in SMS/Email) or 1 (they clicked)
last_step	Format: string. It can have one of the following values: "received", "bounced", "saw review", "added to cart", "payment page", "purchased"
nb_units	Format: integer, representing the number of units of customers' order.
order_value	Format: float, representing value of the order customer made. Already minus the coupon applied.

The column 'last step' is the point of contact with customers before they leave our website. Its value are explained below:

- Received: sms/email sent but not opened
- Bounced: opened but did not finished 3 survey questions used to activate promotional coupon.

- Saw review: scrolled down and read information on the product
- Added to cart: customer added product to checkout.
- Payment page: checked out but not purchased
- Purchased made an order.

The 3 survey questions to activate a coupon are as follow:

- What is the price of the last razor you purchased (dollars)
- How long do you typically use a razor before replacing it? (in months)
- How satisfied are you with your current shaving experience? (1: Not satisfied, 5: Very satisfied)

## 7. Identify Response and Exploratory Variables

Response variable:

- Customers' last step ('last\_step')
- Customers' previous price point ('prev\_price')
- Customers' replacement period ('replace\_time')
- Customers' previous satisfaction ('satisfaction\_score')

Exploratory variables:

- Age group ('age\_range')
- Channel Used ('channel')

## 8. Choose Graphical Displays for EDA

Variable (Original & Aggregated)	Choice of Graphical Display
Age	Histogram
Time series of daily conversion percentage for each age group	Line plot
Time series of daily conversion percentage for each channel	Line plot
prev_price	Histogram, Box graph
replace_time	Histogram, Box graph
satisfaction_score	Histogram, Box graph

## 9. Define Analysis Technique

### a. Hypothesis for Response and Turnover Rates across Promotional Channels:

- Test: T-test
- Significance level:  $\alpha = 0.05$
- Steps:
  - Test for normality of response rates for SMS and Email groups.
  - Check for equality of variances using Levene's Test.
  - If assumptions are met, perform the T-test to compare the means of response and conversion rates between SMS and Email.

### b. Hypothesis for Response and Turnover Rates Across Age Groups:

- Test: ANOVA
- Significance level:  $\alpha = 0.05$
- Steps:
  - Test for normality within each age group.
  - Check for homogeneity of variance using Levene's Test.
  - If assumptions are satisfied, perform ANOVA to compare the means of response and conversion rates across age groups.

### c. Hypothesis for Experience Across Age Groups:

- Test: ANOVA
- Significance level:  $\alpha = 0.05$
- Steps:
  - Test for normality within each age group.
  - Check for homogeneity of variance using Levene's Test.
  - If assumptions are satisfied, perform ANOVA to compare the means of response and conversion rates across age groups.