

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ КОРАБЛЕБУДУВАННЯ
імені адмірала Макарова

Факультет економіки моря
(повна назва підрозділу)

КОНСПЕКТ ЛЕКЦІЙ
з навчальної дисципліни
«Аналітика даних у маркетингу»

для здобувачів вищої освіти освітнього ступеня «магістр»
спеціальності 051 «Економіка»

Розробник: д.е.н., проф Дибач І.Л.
Д.е.н., проф. Філіпішина Л.М.

The Big Data Analytics and Business Intelligence lecture notes are developed in the framework of ERASMUS+ CBHE project “Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan” / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.

Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.

Миколаїв
2021

Тема 1. Вступ до аналітики даних в маркетингу та задоволеності клієнтів

Маркетингова аналітика складається з процесів та технологій, які дозволяють маркетологам оцінити успіх та цінність своїх маркетингових ініціатив, визначити тенденції та закономірності та прийняти рішення, що ґрунтуються на даних. Загалом, маркетингова аналітика спрямована на досягнення чотирьох цілей:

- Виміряти ефективність маркетингових кампаній;
- Знайти можливості росту завдяки ефективному маркетингу;
- Розуміти своїх клієнтів;
- Розуміти своїх конкурентів.

Найбільш базовою формою маркетингової аналітики є надання маркетологам інструментів для розуміння того, який вплив на бізнес мають їхні маркетингові кампанії. Це завдання може варіюватися від такого простого, як надання стандартних показників (рейтинг кліків, рентабельність інвестицій тощо) на рівні кампанії, до такого складного аналізу, як розробка моделі ринкового міксу для розробки оптимальної маркетингової стратегії для максимального отримання прибутку.

Заглибившись у демографічні показники та поведінку клієнтів, можна допомогти зрозуміти, які шанси на успіх матимуть різні стратегії. Потім ця інформація може бути використана маркетологами при виборі цільової аудиторії. Завдяки видобутку даних та статистичному моделюванню, маркетингові аналітики можуть забезпечити глибоке розуміння клієнтів і те, що забезпечує успіх.

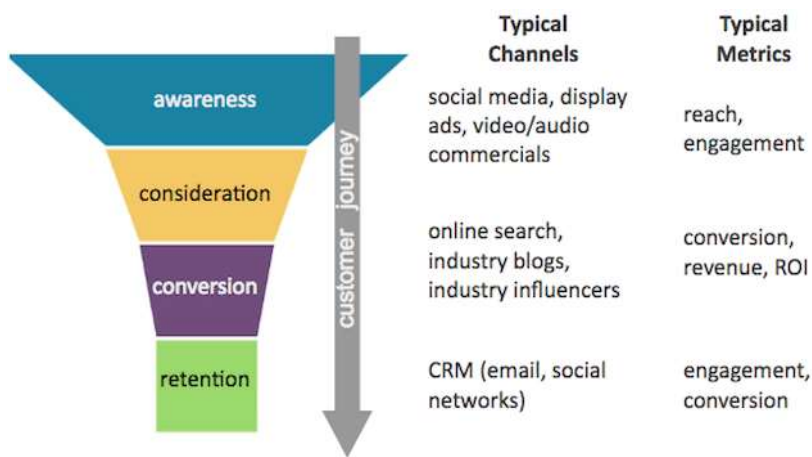
Дослідження ринку часто входять до сфери маркетингової аналітики, і це може допомогти маркетологам краще зрозуміти конкуренцію та відповідно скорегувати свою стратегію.

Як оцінити успіх?

Маркетингова воронка B2C - це зображення типової подорожі клієнта. Вона складається з чотирьох етапів:

1. Усвідомлення
2. Розгляд
3. Перетворення
4. Утримання

Вона візуалізується як послідовність, оскільки очікується, що на вершину націлено багато людей, але все менше і менше людей проходять усі етапи подорожі клієнта з даною компанією.



Усвідомлення

На етапі усвідомлення, який також називають «вершиною послідовності», ви намагаєтесь довести до аудиторії своє існування. Необхідно дізнатись де ваша аудиторія проводить свій час в Інтернеті, і потрібно розміщувати там рекламу - публікації та оголошення в соціальних мережах, рекламні оголошення в Інтернеті, відеорекламу YouTube тощо.

Опосередковано, за допомогою маркетингу поінформованості (також його ще називають рекламою бренду) також намагаються у майбутньому нарощувати трафік із більшою кількістю конверсій. На жаль, неможливо або майже не завжди можливо точно виміряти ці майбутні непрямі наслідки на успішність кампанії.

Якщо кампанія є досить великою та відносно ізольованою від інших змін у бізнесі, можна скористатися Моделюванням ринкових поєднань, щоб отримати оцінку її довгострокового ефекту. Також можна співпрацювати зі стороннім постачальником для вимірювання поінформованості про бренд. Обидва ці методи рідко використовуються невеликими компаніями. Компанії, які інвестують у рекламу бренду, визнають довгострокову цінність сильного бренду, знаючи, що його неможливо легко визначити кількісно.

Основні цілі інформаційних кампаній:

- Збільшити свою органічну аудиторію та залучити її до взаємодії, тому, коли у вас є повідомлення (наприклад, велике оголошення про сезонний розпродаж), воно може бути розширене через вашу мережу.

Метрики: охоплення, частота та залучення.

- Збільшити майбутній органічний трафік та/або збільшити коефіцієнт конверсії майбутніх кампаній із нижчим рівнем послідовності.

Метрики: На жаль, не існує простих методів для кількісної оцінки цієї мети.

Розгляд/конверсія

Коли клієнти активно розглядають можливість здійснення покупки (або іншої конверсії, якщо ви не є веб-сайтом електронної комерції), ваше повідомлення набагато більш адаптоване та транзакційне. Ви знаходите цих клієнтів, коли вони

шукають в Інтернеті ваш продукт (SEO, платний пошук) і досліджують ваш продукт на веб-сайтах експертів (реферали, афілійовані особи, особи, що впливають). Ваше повідомлення більше не має надихати. Ви прагнете переконати та стимулювати перетворення.

Основні цілі розгляду/конверсії кампаній:

- Здійснити продаж (або іншу конверсію).

Показники: дохід (або конверсії), рентабельність інвестицій.

- Запросити клієнта зареєструватися. Хоча першочерговим завданням є закриття продажу, одержання контактної інформації дозволить вам продовжувати намагатися перетворити користувача через постійні, персоналізовані маркетингові відносини.

Метрики: реєстрація на інформаційний бюлетень, завантаження програм.

Утримання

Після того, як клієнт зробив покупку, ви хочете, щоб ваш бренд був у нього на думці наступного разу, коли у нього виникне потреба у вашому продукті. На цьому етапі, ймовірно, у вас є його контактна інформація для надсилання електронних листів, розсилки повідомлень або обміну публікаціями через соціальні медіа. У вас буде купа клієнтів, які ще не готові до наступної покупки (знаходяться високо у воронці), і деякі, які вже готові здійснити таку покупку. В ідеалі, ви можете визначити користувачів, готових до покупки, на етапі конверсії (можливо, за тим, чи нещодавно вони відвідували ваш сайт або переглядали сторінки продуктів), і надсилати їм різні цільові повідомлення.

Основні цілі утримувальних кампаній:

• Тримати аудиторію передплаченою та залученою, щоб, коли вона буде готова до свого наступного циклу покупок, ви були першими про кого клієнти подумують; надихнути на наступний цикл конверсії.

Метрики: охоплення та залучення.

• Отримувати повторні конверсії, особливо клієнтів, які знову перебувають на стадії розгляду.

Метрики: коефіцієнт конверсії, дохід.

Ось кілька найкращих практик використання маркетингової аналітики для розуміння та покращення вашої маркетингової ефективності:

Забезпечте якісні дані. Ваша аналітика спирається на ваші дані. Для середньостатистичного маркетолога дані можуть міститися у понад 17 різних інструментах та додатках. Можливість отримати доступ до всіх цих даних в одному місці, об'єднати дані з кількох джерел та створити відповідні моделі та звіти для всіх типів маркетингових даних - це дуже потужна компетенція для сучасних маркетологів. На жаль, більшість маркетологів все ще покладаються на точкові рішення, надані кожним інструментом у своєму стеку маркетингових технологій, і не мають можливості

аналізувати всі свої дані в одному місці.

Найзручніший спосіб об'єднати всі бізнес-дані в одному місці - скористатися автоматизованим інструментом інтеграції даних, таким як Fivetran, для синхронізації всіх даних у сховищі даних. Домашні варіанти вирішення тієї ж проблеми надзвичайно трудомісткі та схильні до помилок.

Після того, як дані будуть збережені, до них можна підключити такий інструмент бізнес-аналітики, як Chartio, щоб забезпечити візуалізацію, звіти та інформаційні панелі.

Хоча може виникнути спокуса відстежувати якомога більше показників ефективності, краще починати з малого, з кількох варіантів використання та відповідних інформаційних панелей. Тоді ви можете контролювано збільшувати кількість аналізованих показників, переконуючись, що це приносить користь компанії. Для цього визначте свої цілі та виміряйте результати для найбільш важливих для вас випадків використання інструментів маркетингу. Для цього важливо розуміти, які показники мають бути відображені на ваших інформаційних панелях.

Вміло використовуйте аналітику.

Простий метод ґрунтується на швидкому циклі перевірки та навчання, який слід використовувати, щоб допомогти визначити, які показники потрібні. Нова інформація - це можливість дізнатися, прийняти рішення та діяти. Таким чином, кожен раз, коли ми вимірюємо продуктивність, це фактично є можливість розвивати бізнес, а збільшення частоти цього циклу експоненціально впливає на зростання. Спробуйте різні комбінації даних, щоб побачити, що допоможе вам приймати найкращі рішення. Ітерація на ваших інформаційних панелях допомагає вам визначити правильні показники ефективності.

Виберіть відповідні аналітичні візуалізації.

Зацікавлені сторони та маркетингові групи повинні мати можливість зрозуміти маркетингові дані, а візуалізація допомагає це забезпечити. Створюючи діаграми та інформаційні панелі, вибір найбільш підходящої візуалізації даних є важливим для точної інтерпретації цих даних, пошуку аналітичних даних та вжиття заходів. Ви повинні вибрати відповідний інструмент аналізу даних, який дозволяє вам налаштовувати візуалізації замість використання діаграм за замовчуванням для відображення даних.

Розробка моделей, які вимірюють і передбачають.

Аналізу минулого недостатньо в маркетингу. Вам потрібно зробити ще один крок і створити прогнози щодо нових кампаній. Наприклад, окрім звітування про ефективність кампанії, ви можете побудувати аналітичну модель, яка передбачає, які сегменти клієнтів будуть найбільш прибутковими і який вид персональної реклами найкраще підійде для клієнта чи сегмента. Поєднання аналітики та прогнозування може допомогти маркетологам приймати кращі рішення щодо збільшення доходів та більш ефективного використання свого бюджету.

Тема 2. Методи вимірювання та масштабування

Рівні вимірювань у статистиці.

Для здійснення статистичного аналізу даних важливо спочатку зрозуміти змінні та те, що слід вимірювати за допомогою цих змінних. У статистиці існують різні рівні вимірювання, і дані, виміряні за їх допомогою, можна загалом класифікувати на якісні та кількісні.

Спочатку давайте розберемося, що таке змінна. Величина, значення якої змінюється в сукупності і може бути виміряне, називається змінною. Наприклад, розглянемо вибірку зайнятих осіб. Змінні для цієї сукупності населення можуть бути галуззю, місцезнаходженням, статтю, віком, кваліфікацією, типом роботи тощо. Значення змінних буде відрізнятися для кожного працівника.

Наприклад, розрахувати середню погодинну ставку працівника в США практично неможливо. Отже, вибіркочну аудиторію вибирають випадковим чином, щоб вона належним чином представляла більшу сукупність. Потім розраховується середня погодинна ставка цієї вибіркової аудиторії. За допомогою статистичних тестів можна зробити висновок про середню погодинну ставку решти населення.

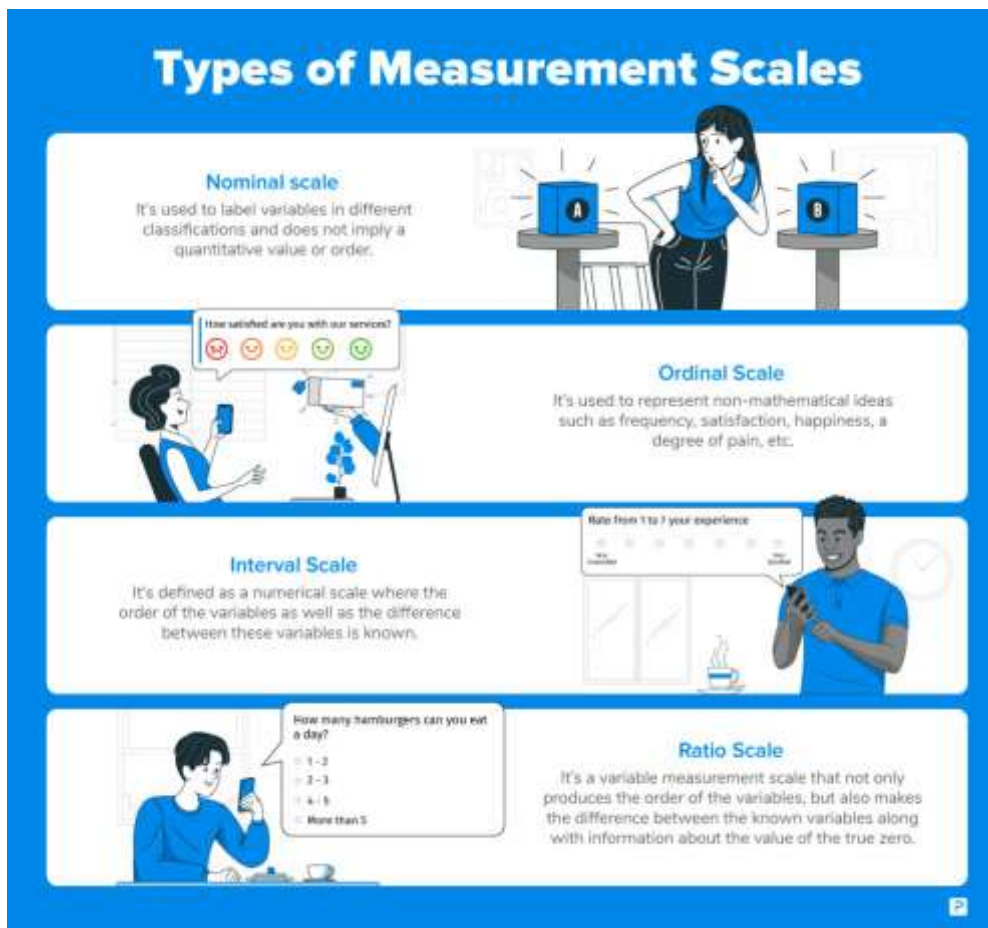
Рівень вимірювання змінної визначає тип статистичного тесту, який буде використовуватися. Математичний характер змінної або іншими словами те, як вимірюється змінна, вважається рівнем вимірювання.

Що таке номінальна, порядкова, інтервальна та співвідносна шкали?

Номінальна, порядкова, інтервальна та співвідносна шкали визначаються як чотири основні рівні шкал вимірювань, які використовуються для збору даних у формі опитувань та анкет, в кожній з яких є питання із кількома варіантами відповіді.

Кожна шкала є поступовим рівнем вимірювання. Перш ніж детально обговорити всі чотири рівні шкал вимірювання з прикладами, давайте коротко розглянемо, що представляють ці шкали.

Номінальна шкала - це масштаб іменування, де змінні просто «називаються» або маркуються без певного порядку. Порядкова шкала має всі свої змінні в певному порядку, окрім простого їх іменування. Інтервальна шкала пропонує мітки, порядок, а також певний інтервал між кожним з варіантів змінної. Шкала співвідношення містить усі характеристики інтервальної шкали, на додаток до цього, вона також може вмістити значення "нуля" на будь-якій з її змінних.



Номинальна шкала: 1-й рівень вимірювання.

Номинальна шкала, також називається категоріальною шкалою змінних, визначається як шкала, що використовується для позначення змінних у окремі класифікації, і не включає кількісне значення чи порядок. Ця шкала є найпростішою з чотирьох шкал вимірювання. Розрахунки, зроблені за цими змінними не проводяться, оскільки вони не містять числових значень.

Бувають випадки, коли цю шкалу використовують для цілей класифікації - числа, пов'язані зі змінними цієї шкали, є лише тегами для категоризації або поділу.

Використовуються для такого питання, як: Де ти мешкаєш?

- 1- Передмістя
- 2- Місто
- 3- Околиця міста

Номинальна шкала часто використовується в дослідницьких опитуваннях та анкетах, де значення мають лише змінні мітки. Наприклад, опитування клієнтів із запитанням "Якій марці смартфонів ви надасте перевагу?" Опціями будуть: "Apple"-1, "Samsung" -2, "OnePlus" -3.

У цьому опитуванні лише назви брендів мають значення для дослідника, який проводить дослідження споживачів. Для цих марок немає необхідності в будь -якому конкретному замовленні. Однак, збираючи номінальні дані, дослідники проводять аналіз на основі відповідних міток.

У наведеному вище прикладі, коли респондент опитування вибирає Apple як бажану марку, введені та пов'язані дані будуть "1". Це допомогло кількісно оцінити та відповісти на останнє запитання - скільки респондентів обрали Apple, скільки вибрали Samsung і скільки вибрали OnePlus - і який з них є найвищим. Це є основою кількісного дослідження, а номінальна шкала - найбільш фундаментальною шкалою дослідження.

Дані номінального масштабу та аналіз.

Існує два основних способи збору даних номінального масштабу:

1. Задаючи відкрите запитання, відповіді на яке можуть бути закодовані до відповідної кількості міток, визначених дослідником.

2. Іншою альтернативою збору номінальних даних є включення запитання з кількома варіантами вибору, у якому відповіді будуть позначені.

В обох випадках аналіз зібраних даних відбуватиметься з використанням відсотків або значень, тобто найбільш поширеної відповіді, отриманої на запитання. Можливо, щоб одне питання мало більше одного значення, оскільки в цільовій сукупності можуть існувати два загальних фаворита.

Приклади номінального масштабу:

- Стать
- Політичні уподобання
- Місце проживання

What is your Gender?	What is your Political preference?	Where do you live?
<ul style="list-style-type: none"> • M- Male • F- Female 	<ul style="list-style-type: none"> • 1- Independent • 2- Democrat • 3- Republican 	<ul style="list-style-type: none"> • 1- Suburbs • 2- City • 3- Town

Порядкова шкала: 2 -й рівень вимірювання

Порядкова шкала визначається як змінна шкала вимірювання, яка використовується для простого відображення порядку змінних, а не різниці між кожною зі змінних. Ці шкали зазвичай використовуються для зображення нематематичних уявлень, таких як частота, задоволеність, щастя, ступінь болю тощо.

Порядковий масштаб зберігає описові якості поряд із внутрішнім порядком, але не має зв'язку із масштабом, і тому відстань між змінними неможливо обчислити. Описові якості вказують на властивості тегування, подібні до номінальної шкали, на додаток до того, порядкова шкала також має відносне положення змінних. Походження цієї шкали відсутнє, через що немає фіксованого початку або "істинного нуля".

Порядкові дані та аналіз.

Дані звичайної шкали можуть бути представлені у табличному або графічному форматі, щоб дослідник міг провести зручний аналіз зібраних даних. Крім того, такі

методи, як U-тест Манна-Уїтні та Тест Крускала-Уолліса Н, також можна використовувати для аналізу порядкових даних. Ці методи зазвичай застосовуються для порівняння двох або більше порядкових груп.

У U-тесті Манна-Уїтні дослідники можуть зробити висновок, яка змінна однієї групи більша чи менша за іншу змінну випадково вибраної групи. У ході тесту Крускала – Уолліса Н дослідники можуть проаналізувати, чи мають дві або більше порядкових груп однакову медіану чи ні.

Приклади порядкової шкали. Статус на робочому місці, рейтинги команд турніру, порядок якості продукції, порядок узгодження чи задоволення - одні з найпоширеніших прикладів порядкової шкали. Ці шкали зазвичай використовуються в дослідженнях ринку для збору та оцінки відносних відгуків про задоволеність продуктом, зміну сприйняття з оновленням продукту тощо.

Наприклад, питання семантичної диференціальної шкали, таке як: Наскільки ви задоволені нашими послугами?

- Дуже незадоволений - 1
- Незадоволений - 2
- Нейтральний - 3
- Задоволений - 4
- Дуже задоволений - 5

Тут порядок змінних має першочергове значення, так само як і маркування. Дуже незадоволені завжди будуть гіршими, ніж незадоволені, а задоволені будуть гіршими, ніж дуже задоволені.

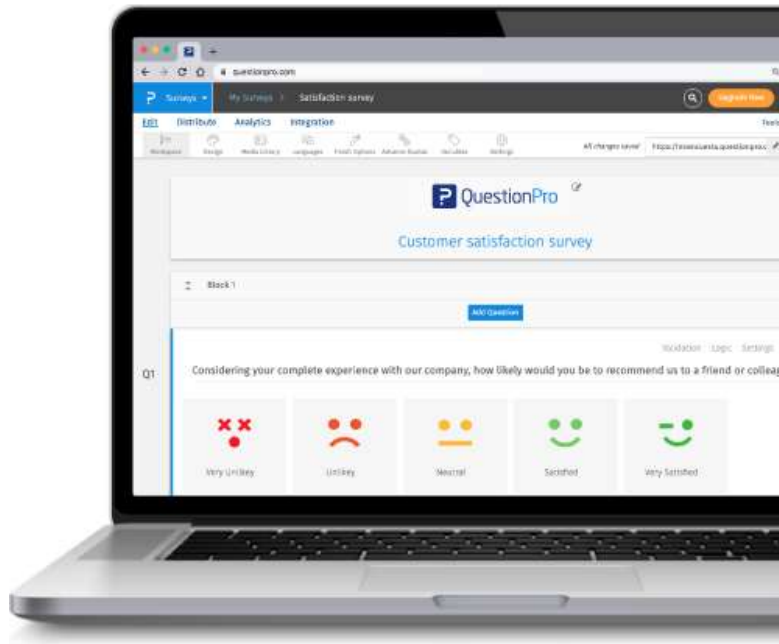
Тут порядкова шкала є на крок вище номінальної шкали - порядок має значення для результатів, а також їх назви.

Аналіз результатів на основі порядку разом з назвою стає зручним процесом для дослідника.

Ця шкала не тільки призначає значення змінним, але також вимірює ранг чи порядок змінних, наприклад: оцінки, задоволення, щастя.

Наскільки ви задоволені нашими послугами?

- 1- Дуже незадоволений
- 2- Незадоволений
- 3- Нейтрально
- 4- Задоволений
- 5- Дуже задоволений



Інтервальна шкала: 3-й рівень вимірювання.

Інтервальна шкала визначається як числова шкала, де відомий порядок змінних, а також різниця між цими змінними. Змінні, які мають відомі, постійні та обчислювані відмінності, класифікуються за шкалою інтервалів. Легко запам'ятати і первинну роль цієї шкали: «Інтервал» вказує на «відстань між двома сутностями», чого допомагає досягти шкала інтервалів.

Ці шкали ефективні, оскільки відкривають двері для статистичного аналізу наданих даних. Для розрахунку центральної тенденції в цій шкалі можна використовувати середнє значення, медіану або значення. Єдиний недолік цієї шкали - відсутність заздалегідь визначеної вихідної точки або справжнього нульового значення.

Інтервальна шкала містить усі властивості порядкової шкали, на додаток до якої вона пропонує розрахунок різниці між змінними. Основною характеристикою цієї шкали є рівновіддалена різниця між об'єктами.

Наприклад, розглянемо шкалу температури Цельсія/Фаренгейта:

- 80 градусів завжди вище 50 градусів, і різниця між цими двома температурами така ж, як і різниця між 70 і 40 градусами.
- Крім того, значення 0 є довільним, оскільки негативні значення температури все-таки існують - що робить шкалу температури Цельсія/Фаренгейта класичним прикладом інтервальної шкали.
- Інтервальну шкалу часто обирають у тих дослідницьких випадках, коли різниця між змінними полягає в інтервалах - чого неможливо досягти за допомогою номінальної або порядкової шкали. Шкала інтервалів кількісно визначає різницю між двома змінними, тоді як інші дві шкали здатні виключно пов'язувати якісні значення

зі змінними.

- Середнє та медіанне значення в порядковій шкалі можна оцінити, на відміну від двох попередніх шкал.

- У статистиці інтервальна шкала часто використовується як числове значення, яке можна не тільки призначати змінним, але також можна проводити розрахунок на основі цих значень.

Навіть якщо інтервальні шкали загалом універсальні, вони не обчислюють значення “справжнього нуля”.

Інтервальні дані та аналіз.

Усі методи, що застосовуються для аналізу номінальних та порядкових даних, також застосовуються до інтервальних даних. Крім цих методів, існує кілька методів аналізу, таких як описова статистика, кореляційний регресійний аналіз, який широко використовується для аналізу даних інтервалів.

Описова статистика - це термін, який використовується для аналізу числових даних, який допомагає змістовно описати, зобразити або узагальнити дані, а також допомагає обчислити середнє значення, медіану та режим.

Приклади шкали інтервалів:

- Бувають ситуації, коли шкали вважаються інтервальними.
- Окрім температурної шкали, час також є дуже поширеним прикладом інтервальної шкали, оскільки значення вже встановлені, постійні та вимірювані.
- Календарні роки та час також підпадають під цю категорію вимірювальних шкал.
- Шкала Лайкерта, оцінка чистого промютера, семантична диференціальна шкала, біполярна матрична таблиця тощо - найбільш часто використовувані приклади інтервальної шкали.

Наступні питання підпадають під категорію «Інтервальна шкала»:

- Який дохід у вашій родині?
- Яка температура у вашому місті?

Шкала співвідношення: 4-й рівень вимірювання

Масштаб співвідношення визначається як шкала вимірювань змінних, яка не тільки формує порядок змінних, але також робить відомі різниці між змінними разом з інформацією про значення істинного нуля. Він обчислюється, якщо припустити, що змінні мають опцію нуля, різниця між двома змінними однакова і між параметрами існує певний порядок.

З можливістю істинного нуля до змінних можна застосовувати різноманітні методи обчислень та описовий аналіз. На додаток до того, що шкала співвідношення робить усе, що може зробити номінальна, порядкова та інтервальна шкала, вона також може встановити значення абсолютного нуля. Найкращі приклади ваг співвідношення

- це вага і зріст. У маркетингових дослідженнях шкала співвідношення використовується для розрахунку частки ринку, річних продажів, ціни на майбутній продукт, кількості споживачів тощо.

Шкала співвідношення надає найбільш детальну інформацію, оскільки дослідники та статистики можуть обчислити центральну тенденцію, використовуючи такі статистичні методи, як середнє, медіана, мода, а також такі методи, як середнє геометричне, коефіцієнт варіації або середнє гармонічне, також можна використовувати на цій шкалі.

Шкала співвідношення враховує характеристику трьох інших шкал вимірювань змінних, тобто маркування змінних, значущість порядку змінних та обчислювану різницю між змінними (які зазвичай є рівновіддаленими).

Через наявність справжнього нульового значення шкала співвідношення не має від'ємних значень.

Щоб вирішити, коли використовувати шкалу співвідношень, дослідник повинен поспостерігати, чи мають змінні всі характеристики інтервальної шкали разом з наявністю абсолютного нульового значення.

Середнє значення, моду і медіану можна обчислити за допомогою шкали співвідношення.

На фундаментальному рівні дані шкали співвідношення мають кількісний характер, завдяки чому для обчислення даних співвідношень можна використовувати всі методи кількісного аналізу, такі як SWOT, TURF, Cross-tabulation, Conjoint тощо. Хоча деякі методи, такі як SWOT та TURF, аналізуватимуть дані співвідношень таким чином, щоб дослідники могли створювати дорожні карти щодо покращення продуктів чи послуг, а перехресне підведення даних буде корисним для розуміння того, чи будуть нові функції корисними для цільового ринку чи ні.

Наступні питання підпадають під категорію «Шкала співвідношень»:

Який зараз зріст вашої дочки?

- Менше 5 футів.
- 5 футів 1 дюйм - 5 футів 5 дюймів
- 5 футів 6 дюймів- 6 футів
- Більше 6 футів

Яка ваша вага в кілограмах?

- Менше 50 кілограмів
- 51-70 кілограмів
- 71-90 кілограмів
- 91-110 кілограмів
- Більше 110 кілограмів

Чотири шкали вимірювання даних - номінальна, порядкова, інтервальна та співвідношення - досить часто обговорюються в академічному навчанні. Нижче наведена діаграма, яка легко запам'ятовується, може допомогти вам у вашому тесті

статистики.

Інші:	Номінальні	Порядкові	Інтервальні	Співвідношення
Встановлена послідовність змінних	–	Так	Так	Так
Мода	Так	Так	Так	Так
Медіана	–	Так	Так	Так
Середнє	–	–	Так	Так
Різницю між змінними оцінити можна	–	–	Так	Так
Додавання та віднімання змінних	–	–	Так	Так
Множення та поділ змінних	–	–	–	Так
Абсолютний нуль	–	–	–	Так

Тема 3. Експериментальний дизайн

Експерименти використовуються для вивчення причинно -наслідкових зв'язків. Ви маніпулюєте однією або кількома незалежними змінними та вимірюєте їх вплив на одну або кілька залежних змінних.

Експериментальне проектування означає створення набору процедур для систематичної перевірки гіпотези. Хороший експериментальний план вимагає чіткого розуміння системи, яку ви вивчаєте.

Спочатку розглянувши змінні та те, як вони пов'язані (Крок 1), Ви можете зробити прогнози, які є конкретними та перевіреними (Крок 2). Спосіб зміни вашої незалежної змінної (крок 3) визначатиме рівень деталізації та зовнішню валідність ваших результатів. Ваші рішення щодо рандомізації, експериментального контролю та дизайну між суб'єктами (крок 4) визначатимуть внутрішню валідність вашого експерименту.

Крок 1: Визначте свої змінні

Почати слід з конкретного дослідницького питання. Ми будемо працювати з двома прикладами дослідницьких питань, одним із наук про здоров'я та іншим з екології:

Приклад запитання 1: Використання телефону та сон.

Ви хочете знати, як використання телефону перед сном впливає на режим сну. Зокрема, ви запитуєте, як кількість хвилин, які людина користується своїм телефоном перед сном, впливає на кількість годин, які вона спить.

Приклад запитання 2: Температура та дихання ґрунту.

Ви хочете знати, як температура впливає на дихання ґрунту. Зокрема, ви запитуєте, як підвищення температури повітря біля поверхні ґрунту впливає на кількість диоксиду вуглецю (CO₂), що вивільняється з ґрунту.

Щоб перевести ваше дослідницьке питання в експериментальну гіпотезу, вам потрібно визначити основні змінні та зробити прогнози про те, як вони пов'язані.

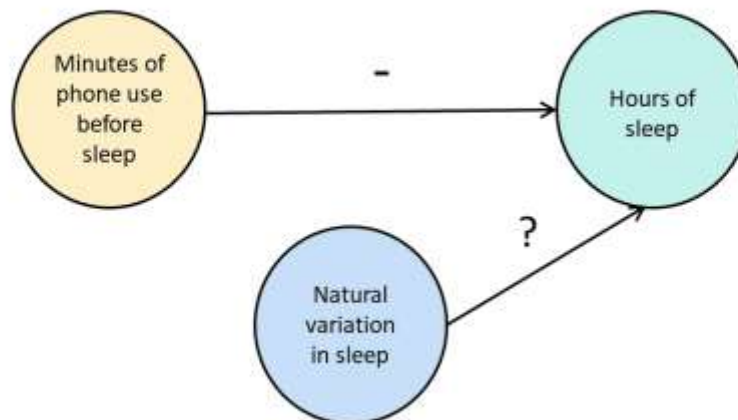
Почніть із простого перерахування незалежних та залежних змінних.

Дослідницьке питання	Незалежна змінна	Залежна змінна
Користування телефоном і сон	Хвилини використання телефону перед сном	Години сну за ніч
Температура і дихання ґрунту	Температура повітря трохи вище поверхні ґрунту	CO ₂ вивільнений з ґрунту

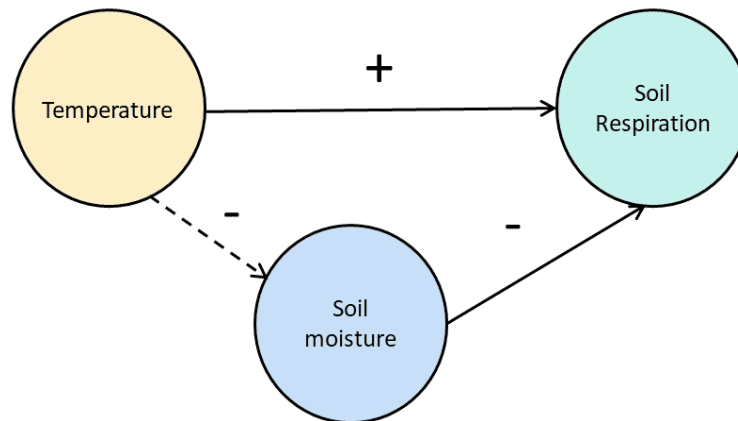
Тоді вам потрібно подумати про можливі сторонні та заплутані змінні та подумати, як ви можете керувати ними у своєму експерименті.

	Стороння змінна	Як контролювати
Користування телефоном і сон	Природні зміни у режимі сну серед людей.	Статистичний контроль: вимірюйте середню різницю між сном під час користування телефоном та сном без використання телефону, а не середньою кількістю сну на групу вимірювання.
Температура і дихання ґрунту	Вологість ґрунту також впливає на дихання, і вологість може зменшуватися підвищенням температури.	Контролюйте експериментально: стежте за вологістю ґрунту та додайте воду, щоб переконатися, що вологість ґрунту є постійною на всіх ділянках спостереження.

Нарешті, об'єднайте ці змінні в діаграму. Використовуйте стрілки, щоб показати можливі зв'язки між змінними та включити знаки, щоб показати очікуваний напрямок зв'язків.



Тут ми передбачаємо, що кількість хвилин використання телефону матиме негативний вплив на години сну, і прогнозуємо невідомий вплив природних змін на години сну.



Тут ми прогнозуємо, що підвищення температури збільшить дихання ґрунту і зменшить вологість ґрунту, тоді як зниження вологості ґрунту призведе до зменшення дихання ґрунту.

Крок 2: Напишіть свою гіпотезу

Тепер, коли у вас є чітке концептуальне уявлення про систему, яку ви вивчаєте, ви зможете написати конкретну, перевірену гіпотезу, яка відповідає на ваше дослідницьке питання.

	Null (H ₀) hypothesis	Alternate (H _a) hypothesis
Користування телефоном і сон	Користування телефоном перед сном не корелює з тривалістю сну.	Збільшення використання телефону перед сном призводить до зменшення сну.

Наступні кроки описують, як створити контрольований експеримент. У контрольованому експерименті ви повинні вміти:

- Систематично і точно керувати незалежними змінними.
- Точно виміряти залежну змінну (ні).
- Контролювати будь-які потенційно заплутані змінні.

Крок 3: Розробіть експериментальні методи вимірювання

Те, як ви маніпулюєте незалежною змінною, може вплинути на зовнішню валідність експерименту - тобто на те, наскільки результати можна узагальнити та застосувати до широкого світу.

По-перше, вам може знадобитися вирішити, наскільки широко змінювати незалежну змінну.

Дослід з утеплення ґрунту

Підвищити температуру повітря можна:

- трохи вище природного ареалу для вашого регіону.
- у більш широкому діапазоні температур, щоб імітувати майбутнє потепління.
- у граничному діапазоні, що виходить за межі будь-яких можливих природних змін.

По-друге, вам може знадобитися вибрати, наскільки точно змінити незалежну змінну. Іноді цей вибір за вас робить ваша експериментальна система, але часто вам доведеться вирішувати, і це вплине на те, наскільки ви можете зробити висновок із своїх результатів.

Експеримент із використанням телефону

Ви можете вважати використання телефону таким:

- категоріальна змінна: або як двійкова (так/ні) або як рівень коефіцієнта (не користування телефоном, низьке використання телефону, велике використання телефону).
- безперервна змінна (хвилини використання телефону вимірюються щогодини).

Крок 4: Віднесіть своїх суб'єктів до досліджуваних груп

Те, як ви застосовуєте експериментальне дослідження до випробуваних, має вирішальне значення для отримання достовірних та релевантних результатів.

По-перше, вам потрібно врахувати розмір вибірки: скільки осіб буде включено в експеримент? Загалом, чим більше предметів ви включаєте, тим більша статистична сила вашого експерименту визначає, наскільки ви можете довіряти своїм результатам.

Тоді вам потрібно випадковим чином розподілити своїх суб'єктів у групи спостереження. Кожна група отримує різний рівень вимірювання (наприклад, не користується телефоном, мало користується телефоном, сильно користується телефоном).

Слід також включити контрольну групу, яка не зазнає впливу змінних. Контрольна група повідомляє нам, що було б з вашими випробуваними без будь-якого експериментального втручання.

Розподіляючи ваші предмети на групи, необхідно зробити два основні рішення:

1. Повністю рандомізований дизайн проти рандомізованого блочного дизайну.
2. Дизайн між предметами проти внутрішньо-предметного дизайну.

Рандомізація

Експеримент може бути повністю рандомізованим або рандомізованим у межах блоків (страт):

- В повністю рандомізованому варіанті кожен суб'єкт випадково потрапляє до групи дослідження.
- У рандомізованому блочному дизайні (стратифікований випадковий дизайн) суб'єкти спочатку групуються за ознаками спільними для них, а потім випадковим чином розподіляються на спостереження в цих групах.

	Повністю рандомізований дизайн	Випадковий дизайн блоків
Користування телефоном і сон	Усі суб'єкти випадковим чином призначають рівень використання телефону за допомогою генератора випадкових чисел.	Суб'єкти спочатку групуються за віком, а потім у цих групах випадковим чином призначаються процедури користування телефоном.
Температура і дихання ґрунту	Зігрівачі обробки призначаються на ґрунтові ділянки навмання за допомогою генератора чисел для формування координат карти в межах досліджуваної території.	Ґрунти спочатку групуються за середньою кількістю опадів, а потім ділянки обробки розподіляються випадковим чином у межах цих груп.

Іноді рандомізація не є практичною або етичною, тому дослідники створюють частково-випадкові або навіть невивадкові конструкції. Експериментальний план, де спостереження не призначається випадковим чином, називається квазіекспериментальним.

У дизайні між суб'єктами (також відомий як незалежний дизайн заходів або класичний дизайн ANOVA) люди отримують лише один із можливих рівнів експериментального впливу.

У медичних чи соціальних дослідженнях ви також можете використовувати відповідні пари у своєму дизайні між суб'єктами, щоб переконатися, що кожна група містить однакове різноманіття змінних у однакових пропорціях.

У дизайні всередині суб'єктів (також відомі як дизайн повторних вимірювань) кожна особа отримує кожен з експериментальних процедур послідовно, і їх реакція на кожне отримання вимірюється.

Повторні заходи також можуть відноситися до експериментального плану, де ефект з'являється з плином часу, а індивідуальні реакції вимірюються з плином часу, щоб виміряти цей ефект у міру його появи.

Врівноваження (рандомізація або зміна порядку обробки між суб'єктами) часто використовується в конструкціях суб'єктів, щоб гарантувати, що порядок застосування впливу змінних не впливає на результати експерименту.

	Дизайн між предметами (незалежні заходи)	Дизайн всередині предметів (повторні заходи)
Користування телефоном і сон	Піддослідним випадково призначається рівень користування телефоном (низький, середній або високий) і вони дотримуються цього рівня користування телефоном протягом усього експерименту.	Суб'єкти послідовно розподіляються на низький, середній та високий рівень використання телефону протягом усього експерименту, і порядок, у якому вони дотримуються цих процедур, рандомізовано.
Температура і дихання ґрунту	Зігрівачі обробки призначаються на ґрунтові ділянки навмання, і	Кожна ділянка отримує кожен зігрівачу обробку (на 1, 3, 5, 8 і

	Дизайн між предметами (незалежні заходи)	Дизайн всередині предметів (повторні заходи)
	ґрунти зберігаються при цій температурі протягом усього експерименту.	10 ° C вище температури навколишнього середовища) послідовно протягом експерименту, і порядок, у якому вони отримують ці обробки, рандомізується.

Експерименти завжди залежать від контексту, і хороший експериментальний план враховує всі унікальні аспекти вашої навчальної системи для отримання інформації, яка є одночасно дійсною та актуальною для вашого дослідницького питання.

Тема 4. А/В тестування

Тестування А/В дозволяє порівняти коефіцієнти конверсії двох версій сторінки, показуючи половині відвідувачів поточну версію вашого сайту ("Контроль"), а іншу половину змінену сторінку ("Процедура"). Порівнюючи, яка версія отримала більше реєстрацій, продажів або будь-які інші показники, ви можете визначити, яка версія є більш ефективною.

Виконавши А/В тести, ви можете перевірити гіпотезу зі статистично значущими даними експерименту, проведеного з реальними відвідувачами, замість того, щоб приймати неінформовані рішення на основі інтуїції.

Успішні бренди використовують А/В-тестування як постійний процес для керівництва своєю веб-розробкою. Однак після впровадження таких інструментів, як Optimizely та VWO у 2010 році, тестування А/В також стало важливою частиною цифрового маркетингу.



Як працює А/В тестування?

Для того, щоб проводити ефективні тести, вам потрібна «гіпотеза», спосіб редагування вашого сайту та інструмент для запису результатів. Ваша гіпотеза - це просто ваше уявлення про те, як покращити свою веб-сторінку. Це може бути зміна розташування заклику до дії, макет сторінки або навіть колір кнопки.



Програмне забезпечення А/В тестування відстежує та записує вплив змін на поведінку ваших відвідувачів. Він розподіляє трафік між «лікуванням» та

«контролем» та вимірює різні реакції. Більш складні інструменти навіть надсилають більше відвідувачів на сторінку з найефективнішими результатами, тому ви не втратите клієнтів під час запуску тесту.



Як тільки ваш сайт отримає достатньо відвідувань, ваше програмне забезпечення оголосить переможця. Однак перед тим, як зміни можуть бути остаточними, необхідно зробити ще один важливий крок. Аналіз статистичної значущості ваших даних є важливою частиною процесу тестування А/В.

Як і будь-який інший вид тестування, людська помилка часто заважає переконливим результатам. Тож, якщо ви плануєте проводити тести А/В на веб-сайті, переконайтеся, що ви не потрапляєте у дві найпоширеніші помилки:

1. Класична проблема підтвердження власних думок.
2. Намагання зробити висновки без достатнього трафіку, часу чи підвищення (це найпоширеніша помилка тестування А/В)

Хоча більшість інструментів тестування А/В виконає всі розрахунки за вас, все ж варто витратити деякий час, щоб ознайомитися зі статистикою тестування А/В.

А/В тестування проти багатоваріантного - у чому різниця?

Тестування А/В передбачає одну змінну (тобто кнопку із закликом до дії) з двома різними версіями. Якщо тест включає кілька змін, це може бути або "тест А/В/п" або "багатоваріантний тест". На відміну від тесту А/В/п, багатоваріантне (MVT) тестування показує, як різні змінні працюють разом і які з них мають найбільший вплив на ваші конверсії.

Що таке тестування А/В/п?

Тестування кількох версій певного елемента відоме як тестування А/В/п. Припустимо, ви хочете спробувати три різні кольори кнопки, версіями сторінок будуть А, В і С. Оскільки ви можете додати будь-яку кількість різних версій, цей вид тесту стає А, В і "п".



Що таке багатоваріантне тестування?

Багатоваріантне тестування працює так само, але порівнює кілька змінних окремо і в поєднанні. Це дає вам інформацію про те, як працює кожна окрема версія та як різні варіанти працюють разом.



Наприклад, тестування альтернативних версій для двох окремих елементів (кнопка із закликом до дії “X” та зображення заголовка “Y”) означало б порівняння чотирьох комбінацій. Це дасть вам такі комбінації змінних:

- A - Y1 і X1
- B - Y2 і X1
- C - Y1 і X2
- D - Y2 і X2

Багатоваріантне тестування вимагає виключно великого розміру вибірки, тому це дійсно можливо лише для найбільших веб-сайтів.

Що таке роздільне тестування?

Роздільне тестування таке ж, як тестування A/B, за винятком того, що двом сторінкам, A та B, призначаються власні URL адреси. Це прискорює швидкість завантаження сторінок і дозволяє проводити більш масштабні зміни. Однак це також може бути складнішим процесом, і існує більша ймовірність забруднення даних.

Приклади тестування A/B

A/B тестування - це форма «перевірки статистичних гіпотез» - методики, яка виникла на початку 20 -го століття. Такі вчені, як Рональд Фішер, Карл Пірсон та Єжи Нейман, використовували цю техніку у своїх експериментах, встановлюючи такі поняття, як нульова гіпотеза.

У світі маркетингу такі копірайтери, як Клод Хопкінс, застосували ці нові концепції до реклами. Хопкінс використовував коефіцієнт повернення рекламних купонів для вимірювання впливу різних кампаній. Він описав свою техніку в книзі під назвою «Наукова реклама» (1923).

З розвитком Інтернет-реклами та електронної комерції стало можливим автоматизувати тестування на маркетинг та UX -дизайн. Тестування A/B було центральним для успіху незліченної кількості онлайн -бізнесу та зіграло роль у великих історичних подіях.

Приклади успішних A/B тестів

З кінця століття A/B -тестування стало ключовим ресурсом для веб -сайтів SaaS,

електронної комерції та бізнесу. Оскільки легко відбирати та відстежувати відвідувачів веб-сайту, більшість користувачів ніколи не усвідомлюють, що вони є частиною тесту.

- 2000 рік: інженери, які працюють у Google, провели тест, щоб знайти оптимальну кількість результатів для відображення на сторінці результатів пошукової системи. Відповідь (10 результатів на сторінку) залишається відносно послідовною з тих пір.

- 2008 рік: президентська кампанія Барака Обама перевірила ранню сторінку донорів передвиборчої кампанії, визначивши комбінацію копії СТА та зображень, які викликали на 40% більше кліків. Подібні тести призвели до повідомлень про 4 мільйони додаткових реєстрацій та 75 мільйонів доларів додаткових пожертв на кампанії.

- 2009 рік: співробітник Microsoft розробив тип посилання, яке відкривало б сторінки на новій вкладці. Загальна кількість кліків на домашній сторінці MSN збільшилася на 8,9%, що є значним збільшенням залучення користувачів. У тесті взяли участь понад 900 000 британських учасників і повторили (з подібними результатами) у США у червні 2010 року.

- 2009 рік: Google провела випробування більш ніж 40 поєднань кольорів на своїй сторінці результатів пошуку, зупинившись на комбінації, яка, як вважається, приносила їм щорічно додаткові 200 мільйонів доларів.

- 2012: Тест А/В на формат заголовків оголошень приніс 12% додаткового доходу від пошукової системи Microsoft Bing. Це призвело до додаткових 100 мільйонів доларів на рік.

- 2013: Microsoft випробувала альтернативні кольори для заголовків та підписів на своїх сторінках результатів пошуку. Оскільки результати були настільки позитивними, компанія повторила експеримент на більшій вибірці з 32 мільйонів користувачів. Коли були внесені зміни, було показано, що вони приносили додаткові 10 мільйонів доларів доходу щорічно.

Сьогодні такі компанії, як Google, Microsoft, Amazon та Booking, проводять десятки тисяч тестів щорічно. Незважаючи на відносно незначний вплив більшості тестів, вони залишаються найнадійнішим способом покращити продуктивність веб-сайту.

#1 - Тестування заголовків та субтитрів А/В

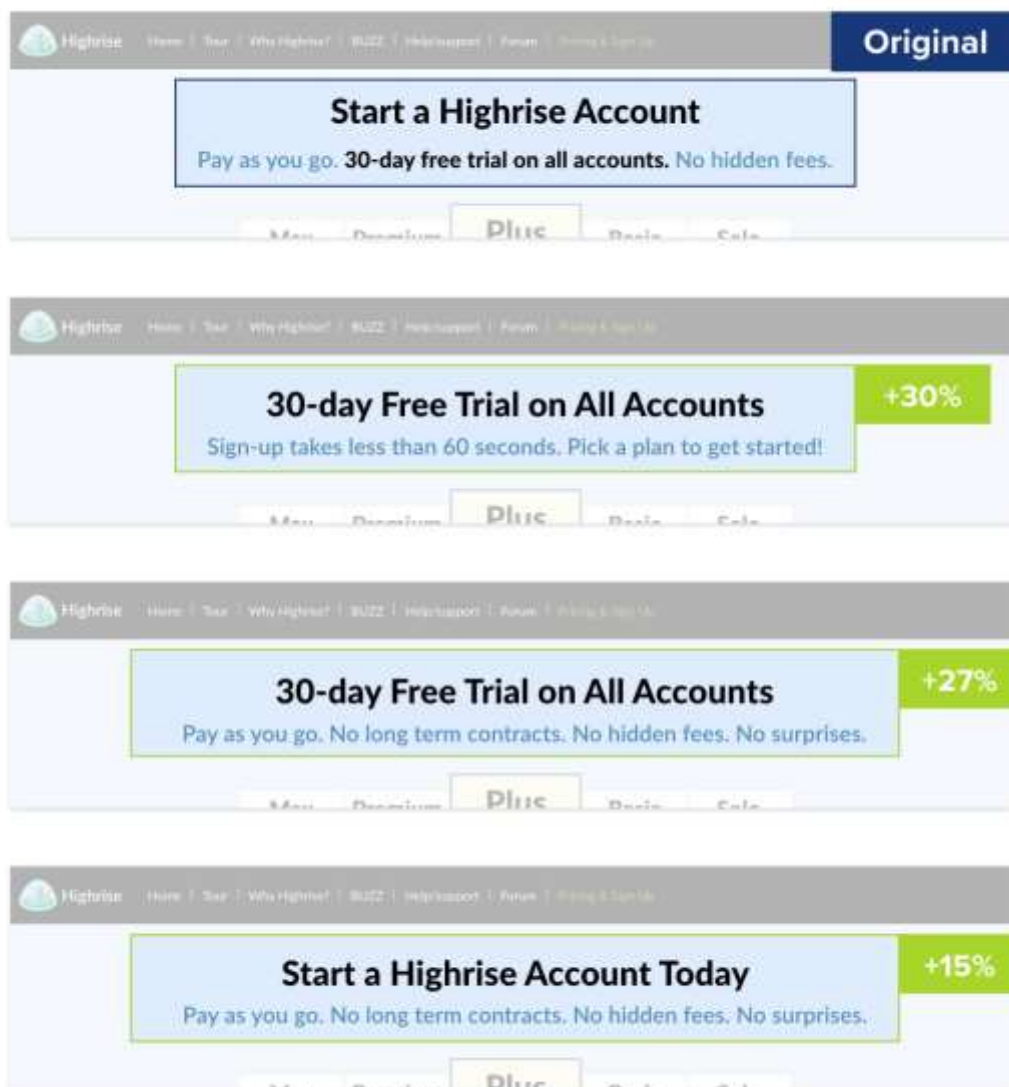
Великий рекламодавець Девід Огілві пояснив важливість сильного заголовка оголошення у своєму «Визнанні рекламної людини»:

В середньому заголовок читають у п'ять разів більше людей, ніж текст. Коли ви написали свій заголовок, ви витратили вісімдесят центів зі свого долара ... Якщо ви не зробили продажу у своєму заголовку, ви витратили 80 відсотків грошей свого клієнта.

Заголовки та субтитри залучають ще більшу частку уваги в Інтернеті, тому має сенс їх перевірити. Насправді, якщо ви не перевірили свій заголовок, ви граєте у

понад 80 відсотків бюджету.

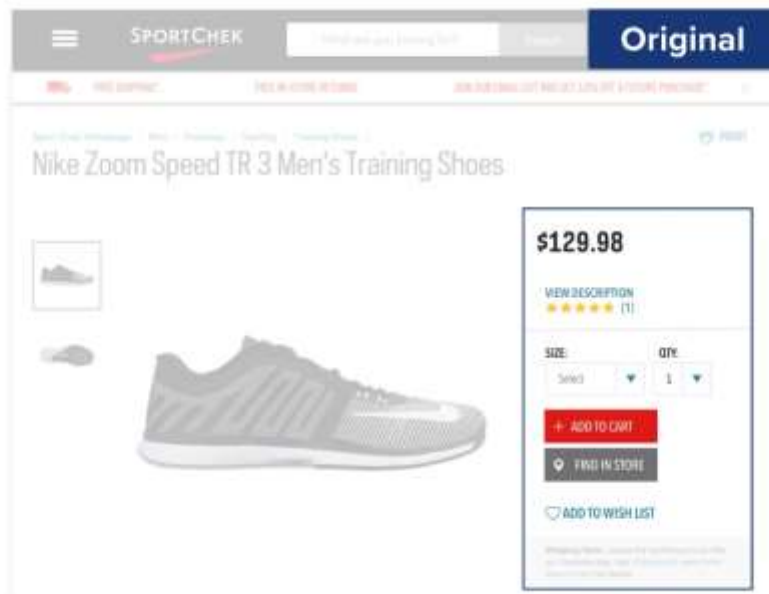
Одним з найвідоміших А/В тестів за останні 10 років була проста сторінка реєстрації для програмного забезпечення CRM Highrise. Тест порівнював чотири різні комбінації заголовків та субтитрів, включаючи оригінал.



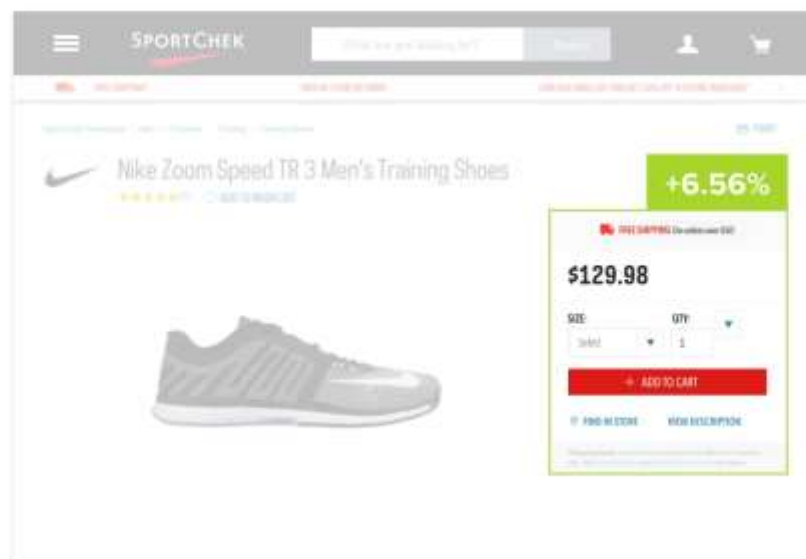
Хоча Highrise використовував невелику вибірку для свого тесту, вони були впевнені в результатах свого експерименту. Заголовок сторінки, який давав тривалість безкоштовної пробної версії та пояснював, наскільки легко було зареєструватися, перетворювався на 30% частіше, ніж оригінал.

#2 - Сповідання веб-сайту тестування А/В

Повідомлення на веб-сайті - це унікальна форма копії продажу. Вони можуть відображати дані в реальному часі або привертати увагу до прихованих деталей, виступаючи в ролі потужного підсилювача вашої ціннісної пропозиції. Як і заголовки та субтитри, сповідання зазвичай приділяють набагато більше уваги, ніж стандартна основна копія.



Відображення деталей доставки на початку процесу оформлення замовлення допомагає зменшити кількість покинутих кошиків. У 2019 році платформа електронної комерції SportCheck експериментувала зі способами рекламувати свою політику безкоштовної доставки на сторінках продуктів та кошиків. Однак результати були невтішними. Додавання сповіщень до сторінок їхніх продуктів збільшило продажі з незначною перевагою, і тест досяг лише 55% впевненості.



Ретельніше вивчивши дані, директор з експериментів виявив, що продажі з тестових сторінок почали падати 29 липня. На цю дату політику магазину щодо безкоштовної доставки було оновлено, і вона застосовувалася лише до замовлень на суму понад 75 доларів США. До оновлення сповіщення збільшило продажі на 6,56% з 96% упевненістю.

#2 - Тестування зображень A/B

Тестування вашого вмісту дозволяє подолати неявні упередження. Наприклад, співробітники президентської кампанії Барака Обами 2007–2008 років виявили, що їх

інстинкти щодо того, які образи найкраще підійдуть, часто були неправильними. Оскільки зображення викликають емоційні та мовчазні реакції, важливо підтвердити свій вибір доказами.



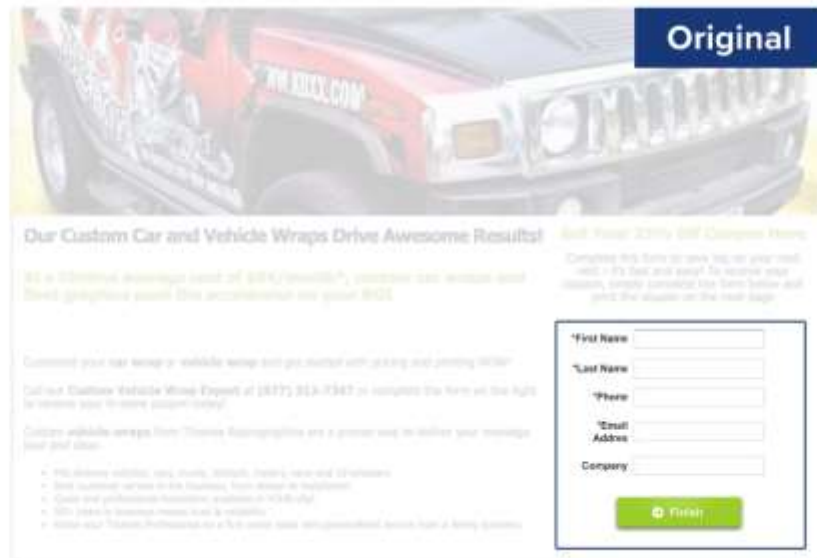
Більшість агентств рекомендують використовувати відео як перше зображення на сторінці товару. Однак у цьому тесті А/В, проведеному продавцем мобільних аксесуарів Zagg, зображення на 360 градусів перевершило відео продукту на 11,9%.



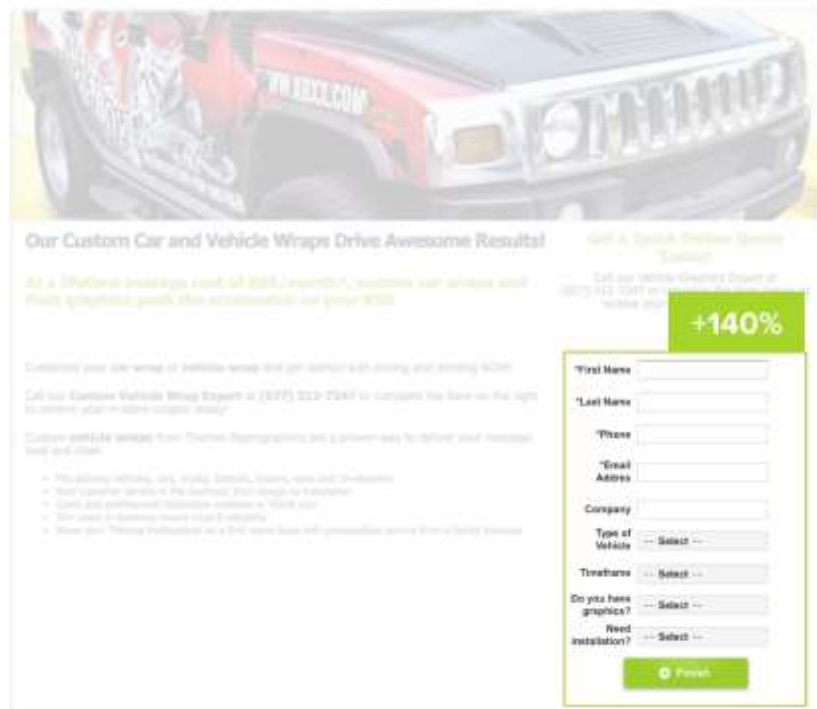
Тест вимірював середній дохід на одного клієнта, так що низькі продажі не могли вплинути на результат. Результати були значними до рівня впевненості 95,4%.

#3 - Тестування веб -форм А/В

Скандально відоме тематичне дослідження "кнопка 300 мільйонів доларів" демонструє цінність тестових форм. Однією з основних перешкод, з якими стикаються веб -дизайнери, є «Прокляття знань». Створити зручний для користувача досвід важко, оскільки дизайнери вже знають, як працюють їх форми та процеси.



Цей тест А/В є ще одним прикладом того, як керівництва з найкращої практики можуть бути помилковими. Фахівці з оптимізації зазвичай наполягають на тому, що коротші форми з меншою кількістю полів, швидше за все, конвертуватимуться ефективніше. Насправді, цей тест від компанії з налаштування автомобілів Thomas Printworks виявив, що довша сторінка з більш конкретними питаннями отримує набагато більш заповнені форми.



Оригінальну сторінку заповнили 2,2% відвідувачів, тоді як довша форма перетворила 5,28% часу. Зростання становило 140% зростання, при рівні впевненості 90%.

Чому тестування А/В настільки популярне у 2021 році?

Тестування А/В має одну велику перевагу перед альтернативними способами

оптимізації веб -сайту: воно базується на реальних користувачах. Хоча дизайн UX, рекомендації щодо найкращої практики та аналіз подорожей клієнтів можуть надати підказки та пропозиції, тестування в реальному світі дає певність.

1. Веб-сайти електронної комерції використовують його для посилення послідовності конверсій

2. Веб-сайти Saas використовують його для покращення своєї домашньої сторінки та покращення процесу реєстрації

3. Веб -сайти покоління потенційних клієнтів використовують його для оптимізації своїх цільових сторінок.

Той самий процес також використовується для зміни дизайну веб -сайтів. Наприклад, у 2017 році British Airways запустила новий веб -сайт. Однак, перш ніж випустити новий дизайн, вони випробували нові версії кожної веб -сторінки. До моменту публікації готового веб -сайту кожна сторінка була перевірена протягом кількох місяців і тисячі відвідувачів.

Як виконати тест А/В?

Тестування А/В пропонує дуже систематичний спосіб дізнатися, що працює, а що не працює в будь -якій маркетинговій кампанії. Більшість маркетингових зусиль спрямовані на збільшення кількості відвідувачів. Оскільки залучення трафіку стає все складнішим та дорожчим, стає першочерговим запропонувати своїм користувачам найкращий досвід відвідування вашого веб -сайту. Це допоможе їм досягти поставлених цілей і дозволить їм здійснити конверсію найшвидшим та найефективнішим способом. А/В тестування в маркетингу дозволяє максимально використовувати наявний трафік і збільшити приплив доходу.

Структурована програма тестування А/В може зробити маркетингові зусилля більш прибутковими, визначивши найважливіші проблемні області, які потребують оптимізації. Тестування А/В тепер переходить від окремої діяльності, яка проводиться раз на блакитному місяці, до більш структурованої та безперервної діяльності, яка завжди повинна проводитися за допомогою чітко визначеного процесу CRO. Загалом він включає наступні кроки:

Крок 1: Дослідження

Перш ніж будувати план тестування А/В, необхідно провести ретельне дослідження того, наскільки веб -сайт працює зараз. Вам доведеться збирати дані про все, що стосується того, скільки користувачів заходить на сайт, які сторінки приносять найбільший трафік, різні цілі конверсії різних сторінок тощо. Використовувані тут інструменти тестування А/В можуть включати інструменти кількісної аналітики веб -сайтів, такі як Google Analytics, Omniture, Mixpanel тощо. Наприклад, ви можете почати зі списку сторінок, які мають найвищий потенційний дохід або найбільший щоденний трафік. Після цього, можливо, вам захочеться глибше зануритися у якісні аспекти цього трафіку.

Інструменти теплової карти - це провідна технологія, яка визначає, де користувачі витрачають найбільше часу, їх поведінку при прокручуванні тощо. Це може допомогти вам визначити проблемні зони на вашому веб -сайті. Ще одним популярним інструментом для більш глибокого дослідження є опитування користувачів веб -сайтів. Опитування можуть виступати як прямий канал між командою вашого веб-сайту та кінцевим користувачем і часто висвітлюють проблеми, які можуть бути упущені в сукупних даних.

Крім того, якісне уявлення можна отримати з інструментів запису сеансів, які збирають дані про поведінку відвідувачів, що допомагає виявити прогалини в подорожі користувача. Насправді, інструменти запису сеансів у поєднанні з опитуваннями аналізу форм можуть виявити уявлення про те, чому користувачі можуть не заповнювати вашу форму. Це може бути пов'язано з деякими полями, які запитують особисту інформацію або користувачів, або, можливо, занадто довго відмовлялися від ваших форм.

Як ми бачимо, і кількісні, і якісні дослідження можуть допомогти нам підготуватися до наступного кроку процесу, зробивши дієві спостереження для наступних кроків.

Крок 2: Перегляньте та сформулюйте гіпотезу

Наблизьтесь до своїх бізнес-цілей, реєструючи дослідницькі спостереження та створюючи гіпотези на основі даних, спрямовані на збільшення конверсій. Без них ваша тестова кампанія нагадує безцільний компас. Якісні та кількісні інструменти дослідження можуть допомогти вам лише у зборі даних про поведінку відвідувачів. Тепер ви несете відповідальність за аналіз та осмислення цих даних. Найкращий спосіб використати кожен частину зібраних даних-це проаналізувати їх, зробити ретельні спостереження за ними, а потім намалювати веб-сайти та уявлення користувачів для формулювання гіпотез, обґрунтованих даними. Після того, як у вас є готова гіпотеза, перевірте її на основі різних параметрів, таких як те, наскільки ви впевнені, що вона виграє, її вплив на макроцілі та наскільки легко її встановити тощо.

Крок 3: Створіть варіації

Наступним кроком у вашій програмі тестування має бути створення варіації на основі вашої гіпотези, та перевірка її А/В на існуючу версію (контроль). Варіація - це інша версія вашої поточної версії зі змінами, які ви хочете перевірити. Ви можете перевірити декілька варіацій проти елемента керування, щоб побачити, який з них працює найкраще. Створіть варіацію на основі своєї гіпотези про те, що може працювати з точки зору UX. Наприклад, чи достатньо людей не заповнюють форми? У вашій формі занадто багато полів? Чи запитують особисту інформацію? Можливо, ви можете спробувати варіант з коротшою формою або інший варіант, опустивши поля, які запитують особисту інформацію.

Крок 4: Запустіть тест

Перш ніж ми перейдемо до цього кроку, важливо обнулити тип методу

тестування та підхід, який ви хочете використовувати. Після того, як ви заблокували один із цих типів і підходів, які ґрунтуються (див. Вищенаведені глави) на потребах та бізнес-цілях вашого веб-сайту, розпочніть тест і дочекайтеся встановленого часу для досягнення статистично значущих результатів. Майте на увазі одну річ - незалежно від того, який метод ви оберете, ваш метод тестування та статистична точність визначають кінцеві результати.

Наприклад, однією з таких умов є терміни проведення тестової кампанії. Час і тривалість тесту повинні бути точно визначені. Обчисліть тривалість тесту, враховуючи середньодобові та місячні відвідування, приблизний існуючий коефіцієнт конверсії, мінімальне покращення очікуваного коефіцієнта конверсії, кількість варіантів (включаючи контроль), відсоток відвідувачів, включених до тесту, тощо.

Крок 5: Аналізуйте результати та впроваджуйте зміни

Незважаючи на те, що це останній крок у пошуку переможця вашої кампанії, аналіз результатів надзвичайно важливий. Оскільки тестування A/B вимагає постійного збору та аналізу даних, саме на цьому етапі розгортається вся ваша подорож. Після того, як ваш тест закінчиться, проаналізуйте результати тесту, враховуючи такі показники, як збільшення у відсотках, рівень впевненості, прямий та непрямий вплив на інші показники тощо. Після того, як ви розглянули ці цифри, якщо тест пройде успішно, розгорніть виграшний варіант. Якщо тест залишається безрезультатним, зробіть висновки з нього та застосуйте їх у своїх наступних тестах.

Тема 5. Метод One-Way ANOVA

Одностороння ANOVA - для чого цей тест?

Односторонній аналіз дисперсії (ANOVA) використовується для визначення того, чи існують якісь статистично значущі відмінності між засобами трьох або більше незалежних (не пов'язаних між собою) груп. Одностороння ANOVA порівнює середні показники між групами, які вас цікавлять, і визначає, чи якісь із цих засобів статистично суттєво відрізняються один від одного. Зокрема, він перевіряє нульову гіпотезу:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

Де

μ = середнє значення по групі

k = кількість груп.

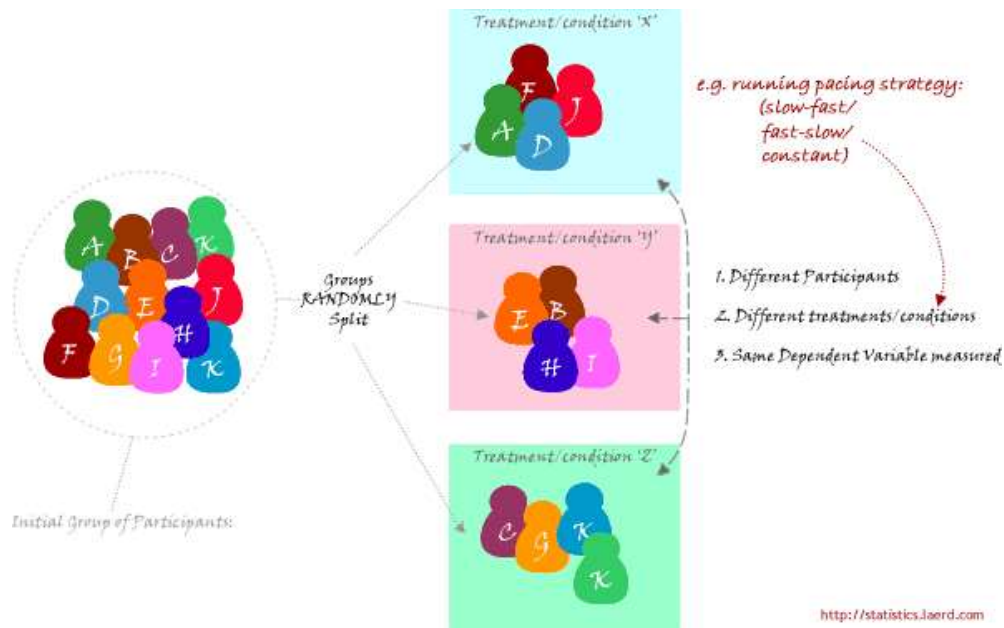
Якщо, однак, одностороння ANOVA дає статистично значущий результат, ми приймаємо альтернативну гіпотезу (H_A), яка полягає в тому, що існує принаймні дві середні групи, які статистично суттєво відрізняються одна від одної.

На цьому етапі важливо усвідомити, що одностороння ANOVA - це універсальна тестова статистика і не може сказати вам, які конкретно групи статистично суттєво відрізнялися одна від одної. Щоб визначити, які конкретні групи відрізняються одна від одної, вам потрібно використовувати пост-хок тест.

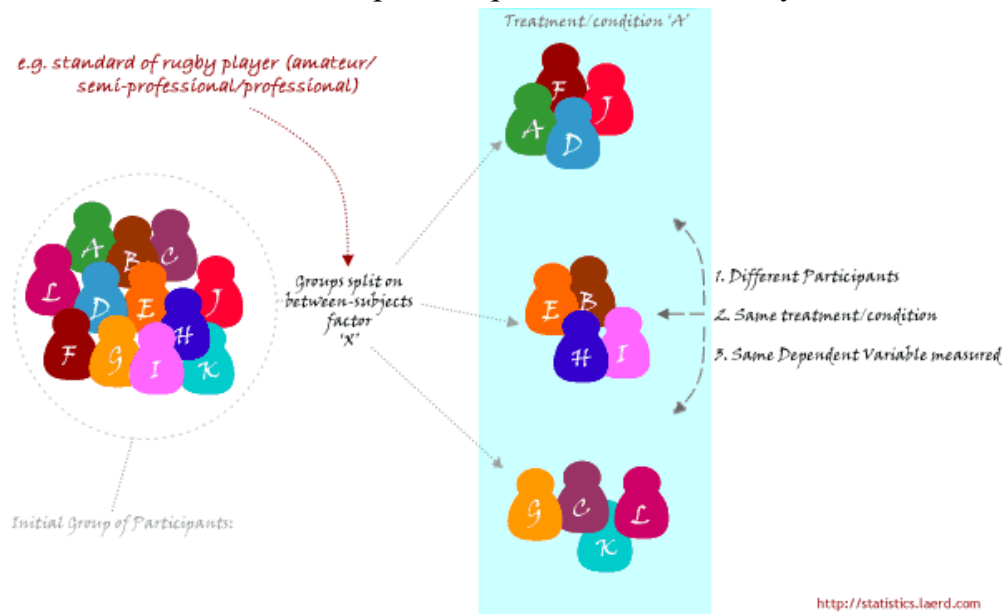
Коли вам може знадобитися скористатися цим тестом?

Якщо ви маєте справу з окремими людьми, ви, ймовірно, зіткнетесь із цією ситуацією використання двох різних типів дизайну дослідження:

Одним із способів дослідження є набір групи осіб, а потім випадковий поділ цієї групи на три або більше менших груп (тобто кожен учасник виділяється в одну і лише одну групу). Потім ви змушуєте кожен групу виконувати різні завдання (або ставити їх в різні умови) та вимірювати результат/відповідь за однією і тією ж залежною змінною. Наприклад, дослідник хоче знати, чи впливають різні стратегії стимуляції на час завершення марафону. Дослідник випадковим чином призначає групу добровольців або до групи, яка (а) починає повільно, а потім збільшує їх швидкість, (б) починає швидко і сповільнює або (в) бігає рівномірно протягом усього. Час завершення марафону є результатом (залежною) змінною. Цей план дослідження схематично проілюстрований на діаграмі нижче:



Другий план дослідження - набрати групу осіб, а потім розділити їх на групи на основі якоїсь незалежної змінної. Знову ж таки, кожна особа буде віднесена лише до однієї групи. Цю незалежну змінну іноді називають змінною, що не залежить від атрибутів, тому що ви поділяєте групу на основі якоїсь ознаки, якою вони володіють (наприклад, їх рівень освіти; кожна особа має рівень освіти). Потім кожну групу вимірюють за однією і тією ж залежною змінною, яка пройшла те саме завдання чи умову (або взагалі жодна). Наприклад, дослідник зацікавлений у визначенні, чи є відмінності в силі ніг між любителями, напівпрофесіоналами та професійними гравцями в регбі. Сила, виміряна на ізокінетичній машині, є залежною змінною. Цей тип дослідження схематично проілюстрований на малюнку нижче:



Чому б не порівняти групи з кількома t-тестами?

Щоразу, коли ви проводите t-тест, є ймовірність, що ви допустите помилку типу I. Ця помилка зазвичай становить 5%. Виконавши два t-тести на одних і тих же даних,

ви збільшите свій шанс "помилитися" до 10%. Формула визначення нової частоти помилок для кількох t-тестів не така проста, як множення 5% на кількість тестів. Однак, якщо ви робите лише кілька багаторазових порівнянь, результати будуть дуже подібними. Таким чином, три t-тести становитимуть 15% (насправді, 14,3%) тощо. Це неприпустимі помилки. ANOVA контролює ці помилки, так що помилка типу I залишається на рівні 5%, і ви можете бути впевненішими, що будь-який статистично значущий результат, який ви знайдете, - це не просто виконання безлічі тестів.

Які припущення робить тест?

Тут є три основні припущення:

1. Залежна змінна зазвичай розподіляється в кожній групі, яку порівнюють в односторонній ANOVA (технічно, нормальні розподіли мають залишки, але результати будуть однаковими). Так, наприклад, якби ми порівнювали три групи (наприклад, аматорські, напівпрофесійні та професійні гравці в регбі) за їх силою ніг, їх значення сили ніг (залежна змінна) мали б бути нормально розподілені для аматорської групи гравців, звичайно розподіляється для напівпрофесіоналів і для професійних гравців.

2. Існує однорідність дисперсій. Це означає, що дисперсії сукупності в кожній групі рівні.

3. Незалежність спостережень. В основному це питання проектування дослідження, і тому вам потрібно буде визначити, чи вважаєте ви, що ваші спостереження можуть бути незалежними на основі вашого плану дослідження (наприклад, робота в групах/сім'ї тощо).

Що станеться, якщо мої дані не виправдають цих припущень?

По-перше, не панікуйте! Перші два з цих припущень легко виправити, навіть якщо останнє припущення - ні. Давайте розглянемо варіанти, описані вище:

1. Одностороння ANOVA вважається надійним випробуванням проти припущення про нормальність. Це означає, що вона досить добре переносить порушення свого припущення про нормальність.

2. Ви можете запустити два тести, які можна застосувати, коли припущення про однорідність дисперсій порушено: (1) Уелч або (2) Тест Брауна та Форсайта.

Тема 6. Метод Two-Way ANOVA

ANOVA (Аналіз відхилення) - це статистичний тест, який використовується для аналізу різниці між середніми показниками більш ніж двох груп.

Двостороння ANOVA використовується для оцінки того, як середнє значення кількісної змінної змінюється відповідно до рівнів двох категоріальних змінних. Використовуйте двосторонню ANOVA, якщо хочете дізнатися, як дві незалежні змінні в поєднанні впливають на залежну змінну.

Приклад: Ви досліджуєте, який тип добрив та щільність посадки дають найбільший урожай у польовому експерименті. Ви призначаєте різні ділянки на полі для поєднання типу добрив (1, 2 або 3) та щільності посіву (1 = низька щільність, 2 = висока щільність), а також вимірюєте кінцеву врожайність культури в гектарах на гектар під час збирання врожаю.

Ви можете скористатися двосторонньою ANOVA, щоб дізнатися, чи впливають тип добрив і щільність посадки на середній урожай.

Коли використовувати двосторонню ANOVA?

Ви можете використовувати двосторонню ANOVA, коли збираєте дані про кількісно залежну змінну на кількох рівнях двох категоріальних незалежних змінних.

Кількісна змінна являє собою кількість. Тонни на гектар є кількісною змінною, оскільки представляє кількість виробленого врожаю. Його можна поділити, щоб знайти середні тонни з гектара.

Категорична змінна представляє типи або категорії речей. Рівень - це окрема категорія всередині категоріальної змінної.

Типи добрив 1, 2 і 3 - це рівні в межах категорійного типу змінних добрив. Густота посадки 1 і 2 - це рівні в межах категоріально змінної густоти посадки.

У вашому наборі даних має бути достатньо спостережень, щоб можна було знайти середнє значення кількісно залежної змінної на кожній комбінації рівнів незалежних змінних.

Обидві ваші незалежні змінні повинні бути категоричними. Якщо одна з ваших незалежних змінних є категоричною, а одна - кількісною, натомість використовуйте ANCOVA.

Як працює тест ANOVA?

ANOVA-тести тестують на значимість, використовуючи F-тест для статистичної значущості. F-тест є груповим порівняльним тестом, що означає, що він порівнює дисперсію у кожній середній групі з загальною дисперсією залежної змінної.

Якщо дисперсія всередині груп менша за дисперсію між групами, F-тест виявить більш високе значення F, а отже, і більшу ймовірність того, що відмінність є реальною, а не випадковою.

Двостороння ANOVA одночасно перевіряє три нульові гіпотези:

- На жодному рівні першої незалежної змінної немає різниці у групових показниках.
- Немає різниці в групових показниках на будь-якому рівні другої незалежної змінної.
- Вплив однієї незалежної змінної не залежить від впливу іншої незалежної змінної (вона ж без ефекту взаємодії).

Двостороння ANOVA без взаємодії (вона ж адитивна двостороння ANOVA) перевіряє лише перші дві з цих гіпотез.

Двосторонні гіпотези ANOVA. У нашому експерименті з урожайності ми можемо перевірити три гіпотези, використовуючи двосторонню ANOVA:

Null hypothesis (H_0)	Alternate hypothesis (H_a)
Різниця в середньому врожаї немає для будь-яких видів добрив.	Існує різниця в середньому врожаї за типом добрив.
Немає різниці в середньому врожаї при густоті посадки.	Існує різниця в середньому врожаї за густотою посадки.
Вплив однієї незалежної змінної на середню врожайність не залежить від впливу іншої незалежної змінної (вона ж без ефекту взаємодії).	Існує різниця в середньому врожаї за типом добрив. Існує різниця в середньому врожаї за густотою посадки. Існує вплив взаємодії між щільністю посадки та типом добрив на середній урожай.

Припущення про двосторонню ANOVA.

Щоб використовувати двосторонню ANOVA, ваші дані повинні відповідати певним припущенням. Двостороння ANOVA робить усі нормальні припущення про параметричний тест різниці:

1. Однорідність дисперсії (вона ж гомоскедастичність).

Різниця навколо середнього значення для кожної порівнюваної групи має бути однаковою для всіх груп. Якщо ваші дані не відповідають цьому припущенню, можливо, ви зможете скористатися непараметричною альтернативою, наприклад тестом Крускала-Уолліса.

2. Незалежність спостережень. Ваші незалежні змінні не повинні залежати одна від одної (тобто одна не повинна викликати іншу). Це неможливо перевірити за допомогою категоріальних змінних - це може бути забезпечено лише за допомогою хорошого експериментального плану.

Крім того, ваша залежна змінна повинна представляти унікальні спостереження - тобто ваші спостереження не повинні групуватися в межах розташування чи окремих осіб.

Якщо ваші дані не відповідають цьому припущенню (тобто якщо ви налаштували експериментальні методи лікування в межах блоків), ви можете включити блокувальну змінну та/або використовувати ANOVA з повторними вимірами.

3. Нормально розподілена залежна змінна. Значення залежної змінної повинні відповідати кривій дзвінка. Якщо ваші дані не відповідають цьому припущенню, ви можете спробувати перетворити дані.

Тема 7. Модель бінарного результату – модель Logit

Що таке логістична регресія?

Логістична регресія - це відповідний регресійний аналіз, який проводиться, коли залежна змінна є дихотомічною (двійковою). Як і всі регресійні аналізи, логістична регресія - це прогнозний аналіз. Логістична регресія використовується для опису даних та пояснення взаємозв'язку між однією залежною двійковою змінною та однією або кількома номінальними, порядковими, інтервальними або незалежними змінними на рівні співвідношення.

Іноді логістичні регресії важко інтерпретувати; інструмент Intellectus Statistics легко дозволяє проводити аналіз, а потім простою англійською інтерпретує результат.

Типи питань, на які може дати відповідь двійкова логістична регресія

Як змінюється ймовірність захворіти на рак легенів (так проти ні) для кожного додаткового фунта людини, який страждає від надмірної ваги, та для кожної пачки сигарет, яку викурюють за день?

Чи впливають маса тіла, споживання калорій, споживання жирів та вік на ймовірність серцевого нападу (так проти ні)?

Основні припущення двійкової логістичної регресії

1. Залежна змінна повинна мати дихотомічний характер (наприклад, наявність проти відсутності).

2. У даних не повинно бути викидів, які можна оцінити шляхом перетворення безперервних предикторів у стандартизовані показники та видалення значень нижче -3,29 або більше 3,29.

3. Серед предикторів не повинно бути високих кореляцій (мультиколінеарність). Це можна оцінити за допомогою кореляційної матриці серед предикторів. Табачник і Фіделл (2013) припускають, що оскільки довгі коефіцієнти кореляції між незалежними змінними менші за 0,90, припущення виконується.

У центрі логістичного регресійного аналізу лежить завдання оцінки журнальних коефіцієнтів події. Математично логістична регресія оцінює функцію множинної лінійної регресії, визначену як:

$$\text{logit}(p) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

for $i = 1 \dots n$.

Надмірна комплектація. При виборі моделі для логістичного регресійного аналізу іншим важливим моментом є відповідність моделі. Додавання незалежних змінних до моделі логістичної регресії завжди збільшить величину дисперсії, поясненої у логарифмічних коефіцієнтах (зазвичай виражається як R^2). Однак додавання все більшої кількості змінних до моделі може призвести до надмірної

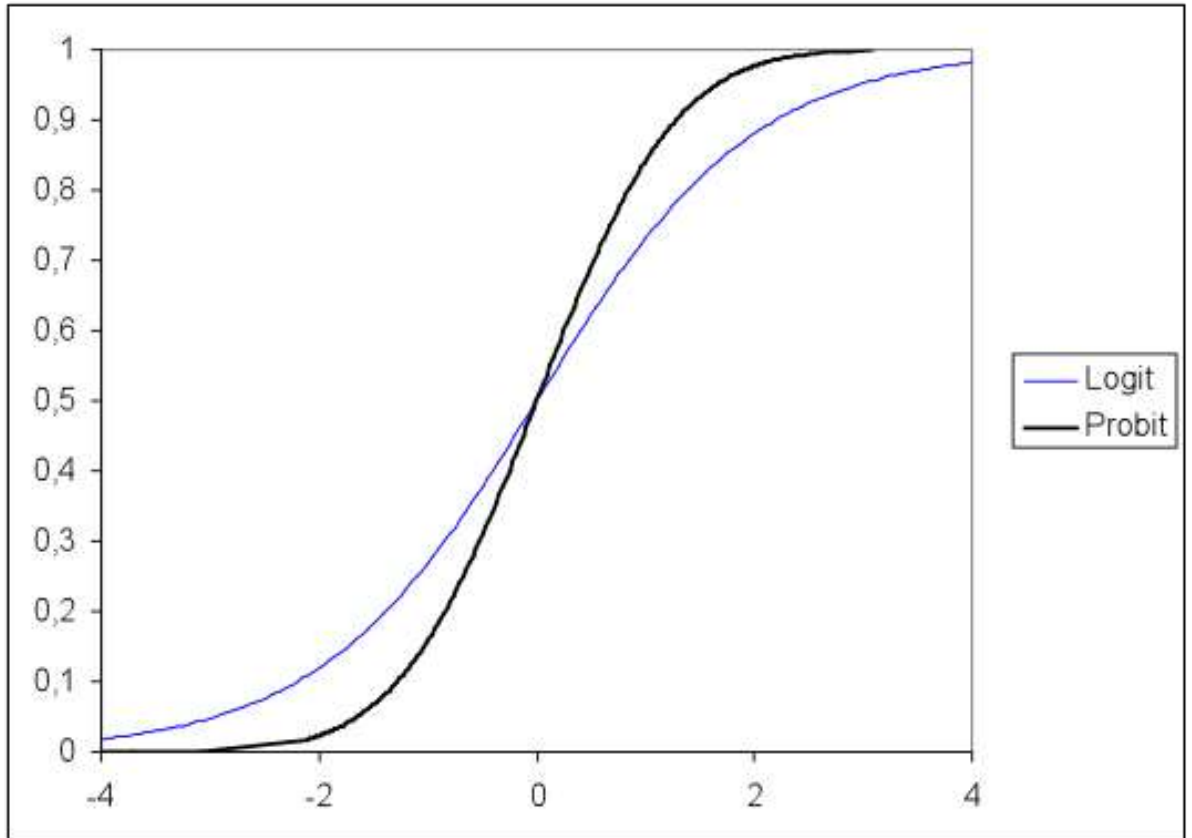
комплектації, що зменшує узагальнення моделі за межами даних, на які вона підходить.

Звіт про R^2 . Для двійкової логістичної регресії було розроблено численні значення псевдо- R^2 . Їх слід тлумачити з особливою обережністю, оскільки вони мають багато обчислювальних проблем, які спричиняють їх штучно високий або низький рівень. Кращим підходом є представлення будь-яких наявних тестів на придатність; Хосмер-Лемешоу-це загальнозживаний показник доброти придатності, заснований на критерії χ^2 -квадрат.

Стандартна лінійна регресія вимагає, щоб залежна змінна мала шкалу безперервного рівня (інтервал або співвідношення). Як ми можемо застосувати той самий принцип до дихотомічної (0/1) змінної? Логістична регресія передбачає, що залежна змінна є стохастичною подією. Наприклад, якщо ми проаналізуємо коефіцієнт загибелі пестицидів, результат буде або вбитий, або живий. Оскільки навіть найбільш стійка помилка може бути лише одним із цих двох станів, логістична регресія передбачає ймовірність зникнення помилки. Якщо ймовірність вбивства клопа більша за 0,5, його вважають мертвим, якщо менше 0,5 - вважають живим.

Досить поширеним є звичайний лінійний регресійний аналіз з фіктивними незалежними змінними. Фіктивна змінна - це двійкова змінна, яка розглядається так, ніби вона буде безперервною. Практично кажучи, фіктивна змінна збільшує перехоплення, тим самим створюючи другу паралельну лінію вище або нижче передбачуваної лінії регресії.

Крім того, ми могли б спробувати просто створити множинну лінійну регресію з фіктивною залежною змінною. Однак цей підхід має два істотні недоліки. По-перше, це може призвести до ймовірностей поза інтервалом (0,1), а по-друге, залишки будуть мати однакову дисперсію.



Щоб вирішити ці недоліки, ми можемо використовувати логістичну функцію, щоб обмежити значення ймовірності до (0,1). The logistic function is $p(x) = 1/(1+\exp(-x))$. Technically this can be resolved to $\ln(p/(1-p)) = a + b \cdot x$. $\ln(p/(1-p))$ is also called the log odds.

Іноді замість логіт -моделі для логістичної регресії використовується модель пробіта. Наступний графік показує різницю для логіт та моделі пробіт для різних значень [-4,4]. Обидві моделі зазвичай використовуються в логістичній регресії; у більшості випадків модель оснащена обома функціями, і вибирається функція, яка найкраще підходить. Однак probit передбачає нормальний розподіл ймовірності події, коли logit припускає розподіл журналу. Таким чином, різниця між логітом і пробітом зазвичай видно лише в невеликих зразках.

У центрі логістичного регресійного аналізу лежить завдання оцінки журнальних коефіцієнтів події. Математично логістична регресія оцінює функцію множинної лінійної регресії, визначену як

$$\text{logit}(p) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 \cdot x_{i2} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{in} \text{ for } i = 1 \dots n.$$

Логістична регресія подібна до дискримінантного аналізу. Дискримінантний аналіз використовує лінію регресії для поділу вибірки на дві групи вздовж рівнів залежної змінної. Якщо в аналізі логістичної регресії використовується концепція ймовірностей і коефіцієнт логарифму з граничною ймовірністю 0,5, то

дискримінаційний аналіз вирізає геометричну площину, представлену хмарою розсіювання. Практична відмінність полягає в припущеннях обох тестів. Якщо дані є багатовимірними нормальними, гомоскедастичність присутня у варіанті та коваріації, а незалежні змінні лінійно пов'язані. Потім використовується дискримінаційний аналіз, оскільки він більш статистично потужний та ефективний. Дискримінаційний аналіз зазвичай більш точний, ніж логістична регресія, з точки зору прогнозної класифікації залежної змінної.

Припущення логістичної регресії

Логістична регресія не робить багатьох ключових припущень щодо лінійної регресії та загальних лінійних моделей, які базуються на звичайних алгоритмах найменших квадратів - особливо щодо лінійності, нормальності, гомоскедастичності та рівня вимірювання.

По -перше, логістична регресія не вимагає лінійного зв'язку між залежними та незалежними змінними. По -друге, значення помилок (залишки) не потрібно нормально розподіляти. По -третьє, гомоскедастичність не потрібна. Нарешті, залежна змінна в логістичній регресії не вимірюється за шкалою інтервалів або співвідношень.

Однак деякі інші припущення все ще діють.

По -перше, двійкова логістична регресія вимагає, щоб залежна змінна була двійковою, а порядкова логістична регресія вимагає, щоб залежна змінна була порядковою.

По -друге, логістична регресія вимагає, щоб спостереження були незалежними один від одного. Іншими словами, спостереження не повинні впливати з повторних вимірювань або відповідних даних.

По -третьє, логістична регресія вимагає, щоб між незалежними змінними була невелика або взагалі відсутня мультиколінеарність. Це означає, що незалежні змінні не повинні надто сильно корелювати між собою.

По -четверте, логістична регресія передбачає лінійність незалежних змінних і коефіцієнт логарифмування. хоча цей аналіз не вимагає лінійного зв'язку залежних та незалежних змінних, він вимагає, щоб незалежні змінні були лінійно пов'язані з коефіцієнтами журналу.

Нарешті, логістична регресія зазвичай вимагає великого розміру вибірки. Загальним орієнтиром є те, що вам потрібно щонайменше 10 випадків з найменш частим результатом для кожної незалежної змінної у вашій моделі. Наприклад, якщо у вас є 5 незалежних змінних і очікувана ймовірність вашого найменш частого результату становить 0,10, то вам знадобиться мінімальний розмір вибірки 500 ($10 \cdot 5 / .10$).

Тема 8. Прогнозування

Успіх багатьох компаній залежить від того, наскільки добре вони прогнозують свій попит. Це один з важливих інструментів для будь-якої організації. Оцінка потреби у майбутньому в теперішній час називається прогнозуванням попиту.

Прогнозування робиться для розуміння попиту, який може виникнути в майбутньому на основі поточних продажів. Ось чому прогнозування часто називають прогнозуванням попиту. Для прогнозування майбутнього враховуються минулі та поточні вимоги.

Крім того, це прогнозування здійснюється науково. Це робиться на подіях. Таким чином, враховуються факти, пов'язані з цим. Тому можна скласти оцінку для прогнозування майбутнього попиту. Це робиться шляхом збору даних про різні аспекти ринку. Таким чином, це концепція, яку ми використовуємо для прогнозування.

У нинішній ситуації існує необхідність приймати правильні рішення. Тому що ринок дуже конкурентний. Тому стає важливим правильно спланувати бізнес-події. Також це рішення, прийняте менеджерами, ґрунтується на його точності. Таким чином, попит стає важливим для досягнення цілей бізнесу.

Методи прогнозування

Не існує жодного такого методу, який би міг успішно передбачити попит компанії. Отже, деякі методи, які можуть бути корисними для прогнозування, такі:

Метод колективної думки

Цей метод зазвичай використовується продавцем компанії. Вони з його допомогою успішно прогнозують продажі компанії в регіоні. Таким чином, для прогнозування майбутніх продажів розраховуються індивідуальні оцінки. Потім на основі кількох факторів, таких як дизайн товару, ціна продажу, рекламні кампанії тощо, ці вимоги переглядаються.

Основним принципом цього методу є те, що продавець найближче до покупців. Таким чином, вони можуть легко дізнатися причину зміни уподобань клієнта.

Статистичний метод

Це один із важливих методів прогнозування. Крім того, цей метод є надійним, науковим і не містить будь-яких упереджень. Використовується основний статистичний метод. Це регресійний аналіз та метод прогнозування тенденцій.

У регресійному аналізі встановлюється зв'язок між незалежною та залежною змінними. Для прогнозування залежною змінною є кількість, що потребується. У той час як незалежними змінними є ціна товару, дохід тощо. Взаємозв'язок тут вважається лінійним.

У методі прогнозування тенденцій організація використовує минулі дані. Таким чином, у цьому методі використовуються дані, накопичені за минулі продажі. Він розміщений у хронологічному порядку для отримання часового ряду. Таким чином,

зображується минула тенденція і на основі цього прогноуються майбутні тенденції. Також передбачається, що минула тенденція збережеться і в майбутньому.

Опитування за вибором покупця

Якщо попит потрібно прогнозувати на короткий період у рік, найкращий спосіб - запитати у клієнта, що він вважає за краще купувати. Таким чином, кожен потенційний клієнт за цим методом безпосередньо опитується. Існує три способи опитування за цим методом.

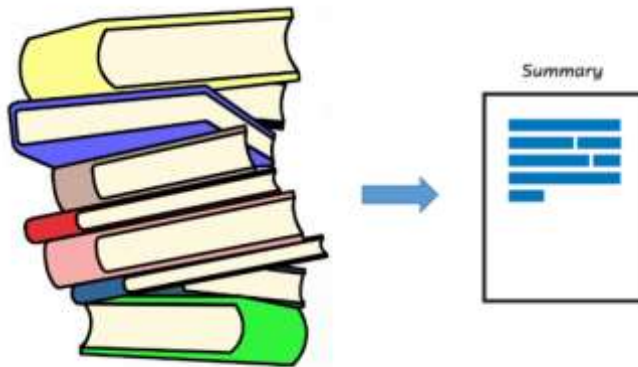
По -перше, ви повинні почати з повного методу перерахування. Тут усіх потенційних клієнтів запитують про плани на майбутнє. Потім слідує вибіркоий метод опитування. При цьому вибірка покупців відбирається науково. Опитуються лише ці люди.

Останній - метод кінцевого використання. Цей метод спеціально використовується для прогнозування попиту на різні ресурси.

Тема 9. Текстове узагальнення

Автоматичне узагальнення тексту привернуло увагу ще в 1950 -х роках. У дослідницькій роботі, опублікованій Гансом Пітером Лунком наприкінці 1950 -х років, під назвою «Автоматичне створення літературних рефератів», використовувалися такі функції, як частота слів та частота фраз, для вилучення важливих речень з тексту для узагальнення.

Інше важливе дослідження, проведене Гарольдом Едмундсоном наприкінці 1960 -х років, використовувало такі методи, як наявність реплік, слів, що використовуються у заголовку, що з'являються у тексті, та розташування речень, для вилучення значних речень для узагальнення тексту. З тих пір було опубліковано багато важливих і захоплюючих досліджень, які вирішують проблему автоматичного узагальнення тексту.

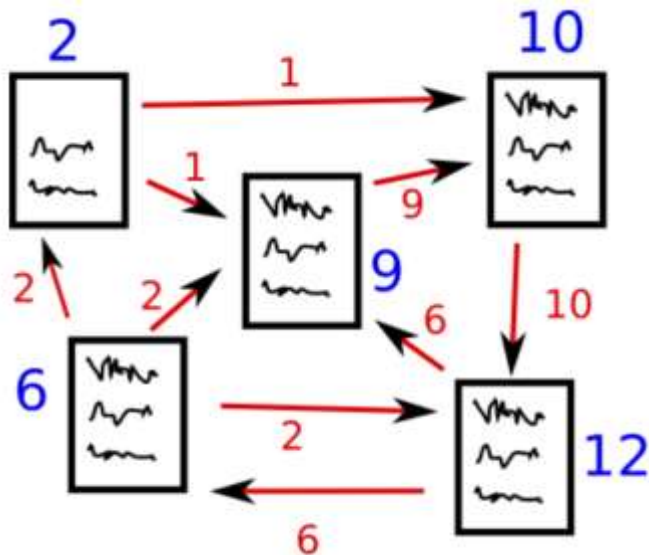


Узагальнення тексту можна загалом умовно поділити на дві категорії - Узагальнююче узагальнення та Абстрактне узагальнення.

1. Узагальнююче узагальнення: Ці методи ґрунтуються на вилученні кількох частин, таких як фрази та речення, з фрагмента тексту та їх складання разом для створення резюме. Тому визначення правильних речень для узагальнення є надзвичайно важливим у методі вилучення.

2. Абстрактне узагальнення: ці методи використовують передові методи НЛП для створення абсолютно нового резюме. Деякі частини цього резюме можуть навіть не відобразитися в оригінальному тексті.

Алгоритм PageRank



Припустимо, у нас є 4 веб -сторінки - w_1 , w_2 , w_3 і w_4 . Ці сторінки містять посилання, що вказують одна на одну. Деякі сторінки можуть не мати посилання - це називаються звисаючими сторінками.

webpage	links
w_1	[w_4 , w_2]
w_2	[w_3 , w_1]
w_3	[]
w_4	[w_1]

- Веб -сторінка w_1 містить посилання на w_2 та w_4
- w_2 містить посилання для w_3 та w_1
- w_4 містить посилання лише на веб -сторінку w_1
- w_3 не має посилань, тому його називатимуть звисаючою сторінкою

Для того, щоб оцінити ці сторінки, нам доведеться обчислити оцінку, яка називається оцінка PageRank. Цей показник - це ймовірність відвідування цієї сторінки користувачем.

Щоб відобразити ймовірність переходу користувачів з однієї сторінки на іншу, ми створимо квадратну матрицю M , що має n рядків і n стовпців, де n - кількість веб -сторінок.

	w1	w2	w3	w4
w1				
w2				
w3				
w4				

Кожен елемент цієї матриці позначає ймовірність переходу користувача з однієї веб -сторінки на іншу. Наприклад, виділена клітинка нижче містить ймовірність переходу від w1 до w2.

	w1	w2	w3	w4
w1				
w2				
w3				
w4				

P(w1 to w2)

Ініціалізація ймовірностей пояснюється наступними кроками:

1. Ймовірність переходу зі сторінки i на j , тобто $M [i] [j]$, ініціалізується $1/(\text{кількість унікальних посилань на веб -сторінці } w_i)$
2. Якщо між сторінками i та j немає зв'язку, то ймовірність буде ініціалізована 0
3. Якщо користувач потрапив на висячу сторінку, то передбачається, що він з однаковою ймовірністю перейде на будь -яку сторінку. Отже, $M [i] [j]$ буде ініціалізовано $1/(\text{кількість веб -сторінок})$

Отже, у нашому випадку матриця M буде ініціалізована таким чином:

	w1	w2	w3	w4
w1	0	0.5	0	0.5
w2	0.5	0	0.5	0
w3	0.25	0.25	0.25	0.25
w4	1	0	0	0

Нарешті, значення в цій матриці будуть оновлюватися ітеративно, щоб потрапити до рейтингу веб -сторінок.

Алгоритм TextRank

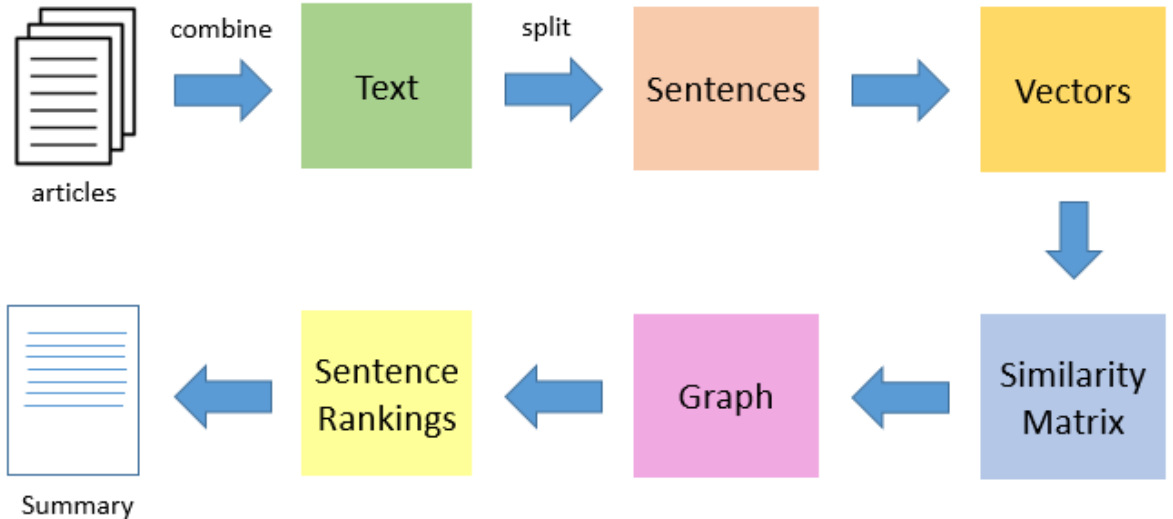
Давайте зрозуміємо алгоритм TextRank, тепер, коли ми зрозуміли PageRank. Нижче перераховані подібності між цими двома алгоритмами:

- Замість веб -сторінок ми використовуємо речення
- Подібність між будь -якими двома реченнями використовується як еквівалент

ймовірності переходу на веб -сторінку

- Оцінки подібності зберігаються у квадратній матриці, подібній до матриці M, що використовується для PageRank

TextRank - це екстрактивна та неконтрольована техніка узагальнення тексту. Давайте поглянемо на алгоритм TextRank, якого ми будемо дотримуватися:



- Першим кроком буде об'єднати весь текст, що міститься у статтях
- Потім необхідно розділити текст на окремі речення
- На наступному кроці ми знайдемо векторне зображення (вбудовування слів) для кожного речення
- Потім подібності між векторами речень обчислюються і зберігаються в матриці
- Матриця подібності потім перетворюється на графік з реченнями як вершинами та оцінками подібності як ребрами для розрахунку рангу речень
- Нарешті, певна кількість речень із найвищим рейтингом формує остаточний підсумок

Тема 10. Тематичне моделювання завдань (LDA)

Що таке моделювання тем?

Моделювання тем - це розділ обробки природною мовою без нагляду, який використовується для представлення текстового документа за допомогою кількох тем, що найкраще пояснює основну інформацію у певному документі. Це дещо схоже на кластеризацію. Тепер замість числових ознак у нас є набір слів, які ми хочемо згрупувати таким чином, щоб кожна група представляла тему в документі.

Навіщо потрібне моделювання тем?

Якщо ми подивимось навколо, то побачимо, що величезна кількість текстових даних навколо нас у неструктурованому форматі у вигляді новин, наукових праць, публікацій у соціальних мережах тощо, і нам потрібен спосіб зрозуміти, упорядкувати та позначити ці дані приймати зважені рішення. Моделювання тем використовується в різних додатках, таких як пошук питань про переповнення стека, подібних один до одного, агрегація та аналіз потоків новин, системи рекомендацій тощо. Усі вони зосереджені на пошуку прихованої тематичної структури в тексті, оскільки вважається, що кожен текст, який ми пишемо, будь то твіт, публікація чи науково - дослідна робота, складається з таких тем, як спорт, фізика, аерокосмічна галузь тощо.

Як зробити моделювання теми?

Наразі існує багато способів моделювання тем, але в цій публікації ми будемо обговорювати імовірнісний підхід до моделювання під назвою Латентне розподіл Діріхле (LDA), розроблений професором Девідом М. Блей у 2003 р. Це розширення ймовірнісної латентної семантики. Аналіз (PLSA), розроблений у 1999 році Томасом Гофманом, з дуже невеликою різницею в тому, як вони ставляться до розповсюдження документів. Тож давайте перейдемо безпосередньо до того, як працює LDA.

Приховане виділення Діріхле

Почнемо з розуміння значення кожного слова в заголовку, оскільки я думаю, що воно містить усе, що нам потрібно знати, щоб зрозуміти, як працює LDA.

Латентний: Це стосується всього, що ми апріорі не знаємо і приховано в даних. Тут теми або теми, з яких складається документ, невідомі, але вважається, що вони присутні, оскільки текст створюється на основі цих тем.

Діріхле: Це "розподіл розподілів". Так, ви правильно прочитали. Але що це означає? Давайте подумаємо про це на прикладі. Припустимо, що є машина, яка виробляє гральні кості, і ми можемо контролювати, чи буде машина завжди виробляти гральні кості з однаковою вагою з усіх сторін, чи буде упередження для деяких сторін. Отже, машина, що виробляє гральні кості, є розповсюдженням, оскільки виробляє гральні кості різних типів. Крім того, ми знаємо, що самі гральні кості є розподілом, оскільки ми отримуємо кілька значень, коли кидаємо кубик. Ось

що означає бути розподілом розподілів, і ось що таке Діріхле. Тут, у контексті моделювання тем, Діріхле - це розподіл тем у документах та розподіл слів у темі. На даний момент це може бути не дуже зрозуміло, але це добре, оскільки ми розглянемо це детальніше через деякий час.

Розподіл: Це означає, що як тільки у нас буде Діріхле, ми виділимо теми документам, а слова документа - темам.

Тепер давайте розберемося, як це працює у моделюванні тем.

Нагадуємо, що LDA каже, що кожне слово у кожному документі походить від теми, а тема вибирається з розподілу кожного документа за темами. Отже, ми маємо дві матриці:

1. $\Theta_{td} = P(t | d)$ - це розподіл ймовірностей тем у документах
2. $\Phi_{wt} = P(w | t)$ - розподіл ймовірностей слів у темах

І ми можемо сказати, що ймовірність слова, поданого документом, тобто $P(w | d)$, дорівнює:

$$\sum_{t \in T} p(w|t, d) p(t|d)$$

де T - загальна кількість тем. Припустимо також, що у нашому словниковому запасі є кількість слів W для всіх документів.

Якщо припустити умовну незалежність, ми можемо це сказати

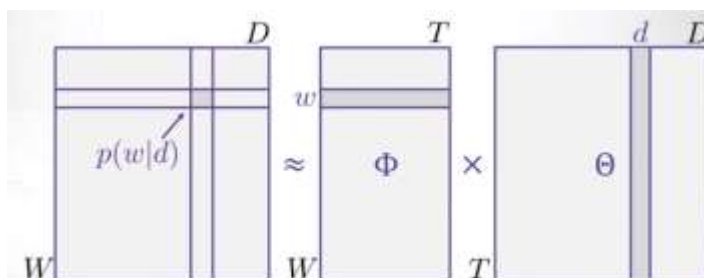
$$P(w|t, d) = P(w|t)$$

And hence $P(w|d)$ is equal to:

$$\sum_{t=1}^T p(w|t) p(t|d)$$

that is the dot product of Θ_{td} and Φ_{wt} for each topic t .

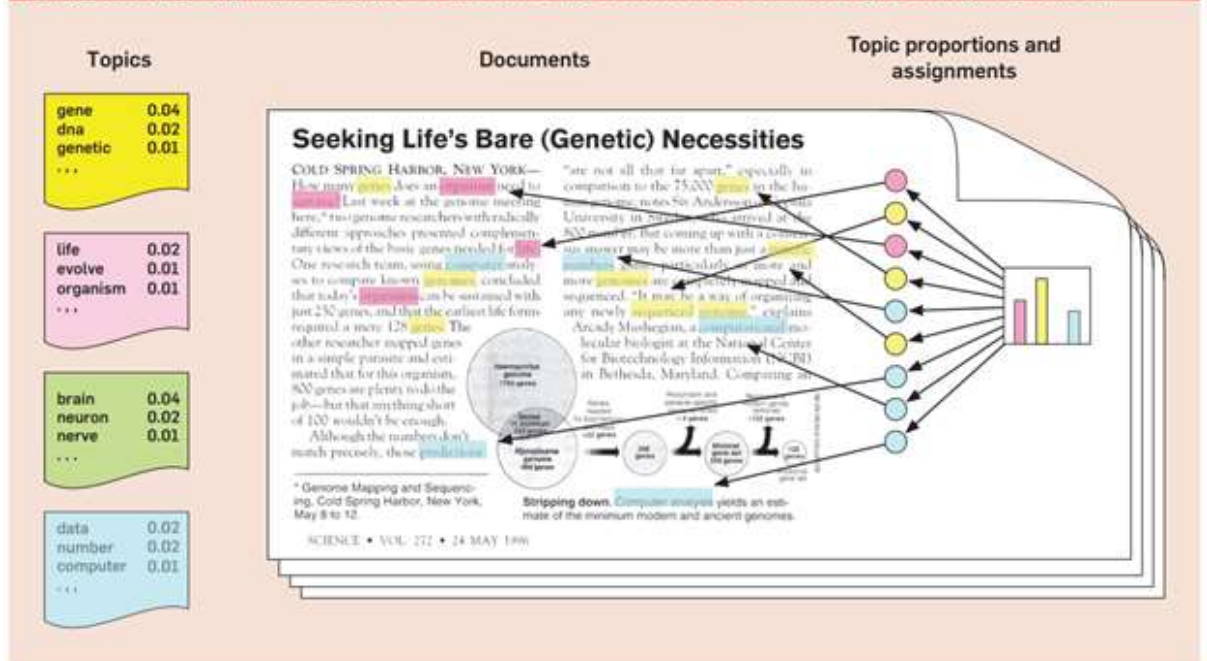
This can be represented in the form of a matrix like this:



Отже, дивлячись на це, ми можемо розглядати LDA, подібно до матричної факторизації або SVD, де ми розкладаємо матрицю розподілу ймовірностей слова в документі на дві матриці, що складаються з розподілу теми в документі та розподілу слів у темі.

Тому ми отримаємо, наприклад, це:

Figure 1. The intuitions behind latent Dirichlet allocation. We assume that some number of "topics," which are distributions over words, exist for the whole collection (far left). Each document is assumed to be generated as follows. First choose a distribution over the topics (the histogram at right); then, for each word, choose a topic assignment (the colored coins) and choose the word from the corresponding topic. The topics and topic assignments in this figure are illustrative—they are not fit from real data. See Figure 2 for topics fit from data.



І щоб зв'язатися з нашим прикладом про гральні кості, ми можемо сказати, що кожне слово у розподілі слів у темі схоже на сторону кості, і у нас є параметр Діріхле для контролю, чи всі слова мають однакову ймовірність у темі або ця тема буде мати крайню упередженість щодо деяких слів. Така ж інтуїція і для розподілу тем у документі.

Добре. Тепер настає важлива частина. Як ми дізнаємось ваги цих двох матриць?

Для початку давайте випадковим чином призначимо ваги обом матрицям і припустимо, що наші дані генеруються згідно з наступними кроками:

1. Випадковим чином виберіть тему з розподілу тем у документі, виходячи з призначених їм ваг. У попередньому прикладі, припустимо, ми вибрали рожеву тему
2. Далі, виходячи з розподілу слів для обраної теми, виберіть слово випадковим чином і помістіть його в документ
3. Повторіть цей крок для всього документа

У цьому процесі, якщо наше припущення про ваги є неправильним, то фактичні дані, які ми спостерігаємо, будуть малоймовірними за наших передбачуваних ваг та процесу генерування даних. Наприклад, припустимо, що у нас є документ D1, який складається з такого тексту:

"Підсистема візуальної обробки Qualcomm® Adreno™ 630 із функцією 6DoF кімнати з SLAM, Adreno Foveation"

і припустимо, що ми присвоюємо високі ваги темі T1, яка має велику вагу для таких слів, як ложка, тарілки, цибуля тощо належить T1. Тому, ми робимо те, що ми намагаємось максимально збільшити ймовірність наших даних, враховуючи ці дві

матриці.

Щоб визначити правильні ваги, ми будемо використовувати алгоритм під назвою вибірки Гіббса. Давайте тепер розберемося, що таке вибірка Гіббса і як вона працює в LDA.

Тема 11. Аналіз частот згадувань (N-gram Frequency Count)

Статистичні мовні моделі, по суті, є типом моделей, які призначають ймовірності послідовностям слів. Ми розберемо найпростішу модель, яка призначає ймовірності реченням і послідовності слів, N-gram.

Ви можете уявити N-gram як послідовність з N слів. За цим поняттям 2-gram (або біграма) - це послідовність із двох слів, таких як «будь-ласка, поверніть», «поверніть своє» або «ваше завдання», А 3-gram (або триграма) - це послідовність слів із трьох слів, наприклад, «будь-ласка, поверніть своє» або «переверніть домашнє завдання»

Модель Bigram

Як випливає з назви, модель bigram апроксимує ймовірність слова з урахуванням усіх попередніх слів, використовуючи лише умовну ймовірність одного попереднього слова. Іншими словами, ви наближаєте це з ймовірністю: $P(\text{те} \mid \text{що})$

І тому, коли ви використовуєте модель bigram для прогнозування умовної ймовірності наступного слова, ви таким чином робите наступне припущення. Це припущення, що ймовірність слова залежить тільки від попереднього слова, також відоме як припущення Маркова.

Моделі Маркова - це клас імовірнісних моделей, які передбачають, що ми можемо передбачити ймовірність якоїсь майбутньої одиниці, не заглядаючи занадто далеко в минуле.

Ви також можете узагальнити модель bigram на модель триграми, яка дивиться на два слова в минуле і, таким чином, може бути додатково узагальнена на модель N-грам

Оцінка ймовірності

Тепер, коли ми розуміємо основну основу моделей N-gram, ви могли б подумати, як ми можемо оцінити функцію ймовірності. Одним із найпростіших та інтуїтивно зрозумілих способів є оцінка максимальної ймовірності (MLE)

Наприклад, щоб обчислити конкретну біграмну ймовірність слова y з урахуванням попереднього слова x , можна визначити кількість біграмів $C(x, y)$ і нормалізувати його за сумою всіх біграм, які мають одне й те саме перше слово x .

Звичайно, є проблеми, як і з кожним підходом до моделювання та методом оцінки. Давайте розглянемо ключові з них, що впливають на модель N-грам, а також використання MLE

Чутливість до навчального корпусу

Модель N-грам, як і багато статистичних моделей, значно залежить від навчального корпусу. В результаті ймовірності часто кодують окремі факти про даний навчальний корпус. Крім того, продуктивність моделі N-грам змінюється зі зміною

значення N .

Більш того, у вас може бути мовне завдання, в якому ви знаєте всі слова, які можуть зустрічатися, а отже, ми заздалегідь знаємо словниковий запас V . Припущення про закритий словник передбачає відсутність невідомих слів, що малоймовірно в практичних сценаріях.

Згладжування

Помітною проблемою підходу MLE є розріджені дані. Це означає, що будь-який N -грам, який з'явився достатню кількість разів, міг би мати розумну оцінку своєї ймовірності. Але оскільки будь-який корпус обмежений, у ньому обов'язково бракує деяких цілком прийнятних послідовностей англійських слів.

В результаті цього матриця N -грамів для будь-якого навчального корпусу обов'язково матиме значну кількість випадків передбачуваних “нульових ймовірностей N -грамів”

Тема 12. Семантичні згадування (Semantic Topic Tagging)

Семантичне ядро сайту - це список ключових слів, які розподілені на сторінках сайту і за допомогою яких відбувається просування сайту в мережі, це слова та фрази, які найточніше описують вид діяльності, товари або послуги, що пропонує певна веб-сторінка. Така серцевина зазвичай складається з високочастотного ключового слова та інших, які ранжуються за частотою використання. Ядро є основою сайту, від нього залежить, чи буде веб-сторінка в ТОП видачі пошукових систем.

Використання семантичного ядра

- 1 За допомогою нього формується тематика сайту
- 2 Є базою для побудови структури ресурсу
- 3 Допомогає ранжувати веб-сторінку в пошукових системах
- 4 Допомогає задовольняти потреби користувачів

Алгоритм створення семантичного ядра для інтернет-магазину складається з 5 кроків:

1. Зібрати ключові слова за допомогою звичайних сервісів та програм
2. Використати альтернативні способи зі збору бази семантичного ядра.
3. Згрупувати дані та видалити зайві ключові слова.
4. Розширити семантику на сторінці.

Що таке ключові слова?

Ключові слова - окремі слова та цілі фрази, за допомогою яких користувачі мережі шукають потрібну їм інформацію. Ключі бувають: Низькочастотні. За різними даними, до таких належить 65-80% всіх пошукових запитів. Тому важливо скласти максимально повну семантику, яка повинна містити багато низькочастотних ключових слів і бути максимально заточеною під запити користувачів. Середньочастотні - до 1000 показів. Потрапляння в ТОП за такими ключами - запорука збільшення трафіку. Високочастотні - від 1000 показів. Рівень конкуренції запитів найвищий.

Види ключових запитів:

Інформаційні. Використовуються аудиторією під час пошуку певних відомостей: "як підключити роутер", "як зберігати печиво", "як написати реферат".

Транзакційні. Використовуються під час планування певних дій: "купити смартфон", "завантажити книгу з лідогенерації", "замовити торт з самовивозом".

Невизначені. Коли складно встановити конкретні наміри автора повідомлення: ключ "торт" можна використовувати перед запланованою покупкою або ж під час пошуку інформації, як приготувати кондитерський виріб самостійно.

Навігаційні. Коли користувач шукає інформацію на конкретній сторінці:

"ноутбук Леново", "Нова пошта відстежити доставку", "реєстрація в Фейсбук".

Кластеризація семантичного ядра

Це процес систематизації пов'язаних між собою ключових слів, які пропонує пошукова видача мережі, в групи (кластери). Це дозволяє сформувати впорядковану структуру з сотень та тисяч пошукових запитів користувачів мережі. Питання групування семантики вимагає індивідуального підходу для кожного окремого проекту/кампанії/бізнесу. На прийняття рішення впливає безліч факторів: структура сайту, вид бізнесу, бюджет, частотність фраз, гнучкість проекту, основний меседж. Головне - розуміти, чому обрано саме такий вид кластеризації, чи буде він оптимальним і чи досягне потрібного результату. Перед початком збору семантики можна написати список стоп-слів. За їхньою допомогою відсікаються непотрібні фрази. Скоротити кількість нерелевантних запитів допоможе перелік мінус-слів. Якщо правильно сформувати їхній список, можна значно скоротити бюджет витрат.

Виділяють такі етапи кластеризації семантичного ядра:

- 1 Визначення вибірки об'єктів для групування.
- 2 Формування критеріїв для оцінювання об'єктів в списку.
- 3 Дефініція ступенів подібності між розглянутими об'єктами.
- 4 Кластерний аналіз для визначення груп.
- 5 Підбиття підсумків та результатів групування.

Принципи кластеризації семантики:

- За етапом воронки продажів (холодні, теплі, гарячі)
- За інтентами (intention - "намір") - словами, об'єднаними наміром (вартість-калькулятор-ціна)
- За морфологічною близькістю слів (замовлення-замовити)
- За смисловою або семантичною близькістю (групуються синоніми: монтаж-установка)

Як створити структуру Вашого веб-ресурсу

Структура сайту- ієрархічна система сторінок, за допомогою якої можна планувати контент, способи подачі інформації, зручність ресурсу для користувачів та пошукових систем. Для побудови структури та впорядкування семантичного ядра підійде інструмент таблиць Word, Excel або таблиці Google. Для цього вкажіть назви сторінок та їхню підпорядкованість, додайте колонки URL сторінок, ключів і частоти їхнього використання.

Сервіси зі збору семантики

У мережі існує чимало різних сервісів зі збору ключових фраз. Використовувати їх можна на початкових етапах та в подальшому - для розширення бази слів. У всіх сервісів різні бази даних, тому вони можуть надавати різні ключові слова. Під час роботи краще використовувати кілька сервісів, звертаючи увагу на такі їхні особливості:

1. Обсяг бази пошукових запитів.
2. Частота оновлення даних.
3. Підбір синонімів.
4. Джерела пошуку ключових слів.
5. Бази всіма мовами та для всіх країн світу.
6. Можливість знайти пошукові підказки.
7. Відображення важливих частотних ключів.

Помилки під час побудови семантичного ядра

Ключі не відповідають тематиці бізнес-ресурсу. Не визначилися з типом запиту (комерційний або інформаційний). Перевірити це можна так: ввести в пошуку питання і, якщо більша частина веб-ресурсів відгукується з пропозицією "купити-замовити", то це комерційна пропозиція.

Семантичне ядро не включає в себе синоніми ключових слів і фраз.

Дані не були розподілені за групами. Їх потрібно кластеризувати, це важливо для структури сайту.

Ключові слова з помилками.

Зібране ядро занадто маленьке. Ядро має включати в себе якомога більше ключових запитів: від декількох сотень до тисяч.

Були введені порожні пошукові слова.

У список додано лише високочастотні запити. Тоді як саме низькочастотні відразу можуть вивести сайт в топ видачі.