



Data Scientist
at Telecommunication Company

Halo!

Perkenalkan saya **Abdullah Ghifari**.

Sebelumnya



Abdullah Ghifari

<https://www.linkedin.com/in/abdullah-ghifari/>

Statistic

Inferential Statistics



Objektif

Memahami dan mampu melakukan itu A/B Testing, Hypothesis Testing, Analysis of Variance beserta prinsipnya dan kapan menerapkannya

Expected Output

Memahami apa itu A/B Testing dan prinsipnya

Mampu menerapkan A/B Testing dan kapan menerapkannya

Memahami apa itu Hypothesis Testing beserta prinsipnya

Mampu menerapkan Hypothesis Testing dan kapan menerapkannya

Memahami apa itu Analysis of Variance beserta prinsipnya

Mampu melakukan Analysis of Variance dan kapan menerapkannya

Hands-On Required :

Hands - On :

1. **Statistics III - Hypothesis Testing.ipynb**
2. **Statistics III - Experimental Design.ipynb**

Dataset :

1. **ab_data.csv**
2. **Iris.csv**
3. **e-commerce_example_dataset.csv**

**Klik disini untuk mengakses
folder Hands-On dan
Dataset**

Inferential Statistics



Hypothesis Testing

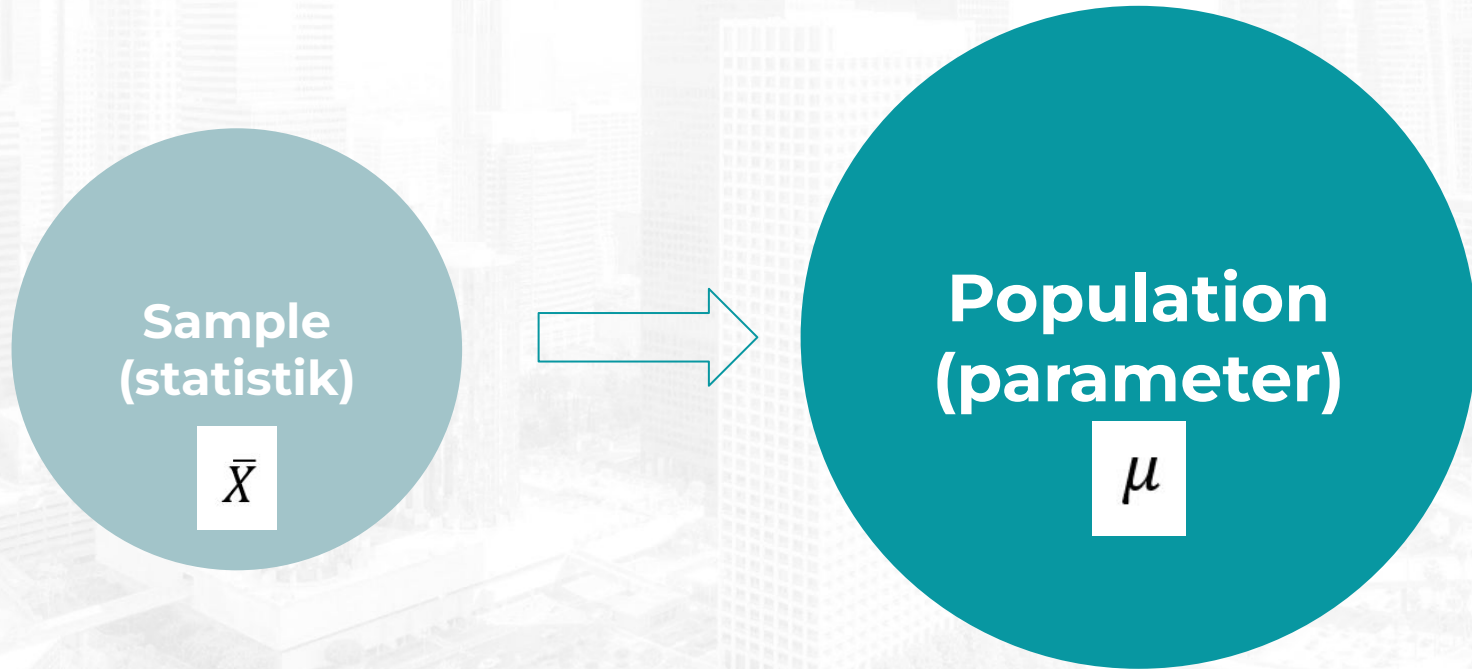


Experimental Design (A/B Testing)



Hands-On

Statistika Inferensi



PT Sukamaju



Rata-rata amount per user : 1.6 M
Standard Deviation : 380k



Rata-rata amount per user: 1.4 M
Standard Deviation : 110k

**Apakah diskonnya mampu
meningkatkan value
penjualan secara signifikan?**

***akan dijawab di sesi hands-on**

The background of the slide is a faded, grayscale image of a city skyline with various skyscrapers and buildings.

Hypothesis Testing (Pengujian Hipotesis)

Tujuan dan Prinsip Pengujian Hipotesis

Tujuan dari pengujian hipotesis adalah membuktikan suatu **pernyataan umum** dengan menggunakan data - data **sampel yang ada**.

Prinsip yang digunakan adalah **praduga tak bersalah** (*Presumption of Innocence*)

Ilustrasi



??????



Ilustrasi



Ali



Hipotesa

Hipotesis 0: Ali **bukan** pencuri

Hipotesis 1: Ali **adalah** pencuri

Ilustrasi

Data-data :

1. Adanya CCTV yang menunjukkan ali di lokasi kejadian
2. Adanya saksi yang menceritakan ali di lokasi tersebut
3. Tertemukannya sidik jari ali di rumah tersebut.

Ilustrasi

Berdasarkan data-data tersebut Ali akan **sangat kuat diduga** sebagai pencuri.

Karena **resiko** untuk salah menduga ali pencuri sangat kecil.

Maka dapat **disimpulkan** bahwa Ali adalah pencuri. (Menerima Hipotesis 1)

Hipotesa

Hipotesis 0: Ali **bukan** pencuri

Hipotesis 1: Ali **adalah** pencuri

Pengujian Hipotesis Dalam Statistika

Hipotesis H_0 yang tadi kita sebutkan dalam statistika biasa disebut dengan Hipotesis Nol / Null Hypothesis (H_0)

Null hypothesis (H_0)

Sebuah hipotesis yang berkontradiksi dari pernyataan yang akan diuji. Sederhananya hipotesis yang bersifat umum.

Pengujian Hipotesis Dalam Statistika

Sedangkan Hipotesis 1 dalam statistika biasa disebut dengan Hipotesis Alternatif / Alternative Hypothesis (H_1/H_a)

Alternative hypothesis (H_1)

Hipotesis yang ingin kita uji. Bersifat spesifik.

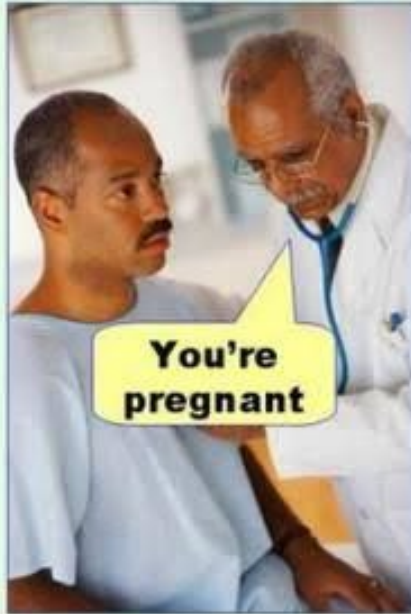
Kemungkinan yang terjadi ketika kita melakukan uji hipotesis

	Kenyataannya Bukan Maling	Kenyataannya Maling
Diduga bukan maling	Divonis bukan maling dan benar	Divonis bukan maling padahal maling (β)
Diduga maling	Divonis maling padahal bukan (α)	Divonis maling dan benar

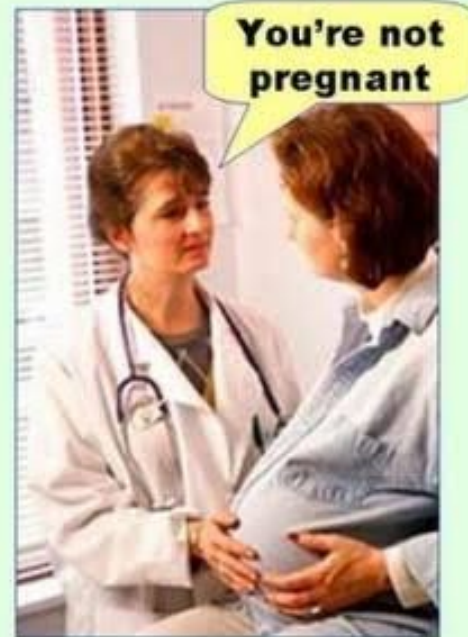
Type I Error : α

Type II Error : β

Type I error
(false positive)



Type II error
(false negative)



Istilah dalam Statistika

	H_0 Benar	H_0 Salah
Menerima H_0	$1 - \alpha$ Tingkat Kepercayaan	β Error Tipe II
Menolak H_0	α Error Tipe I	$1 - \beta$ Kekuatan Uji

α : besar kriteria error tipe I yang bisa kita toleransi

β : besar kriteria error tipe II yang bisa kita toleransi

Metrics untuk mengukur resiko kesalahan

Metrics untuk mengukur resiko kesalahan menolak H_0 (Error Tipe I) dinamakan **P-Value (Probability Value)**.

P-Value ini berbentuk **persentase/peluang** yang menggambarkan resiko ketika kita menolak H_0 /menerima H_1 .

Contoh :

Besarnya resiko kesalahan menyatakan bahwa ali pencuri (p-value) adalah 1%.

Artinya : dari 100 kasus yang serupa akan ada 1 kasus yang salah kita vonis sebagai pencuri.

Cara Mengambil Keputusan

Kondisi pengambilan keputusan dibagi menjadi 2:

1. Jika $P\text{-Value} < \alpha$, maka kita akan mengambil H_1
2. Jika $P\text{-Value} > \alpha$, maka kita akan mengambil H_0

*Besarnya α ditentukan oleh kita atau pertimbangan bisnis (biasanya 5%)

Contoh

Berdasarkan data-data yang diberikan peluang/resiko salah memvonis Ali adalah maling (p-value) adalah 0.03

Contoh

Berdasarkan data-data yang diberikan peluang/resiko salah memvonis Ali adalah maling (p-value) adalah 0.03

Kriterianya $\alpha = 0.05$

Contoh

Berdasarkan data-data yang diberikan peluang/resiko salah memvonis Ali adalah maling (p-value) adalah 0.03

Kriterianya $\alpha = 0.05$

Karena p-value < α , maka

Kesimpulan :

Cukup bukti menyatakan bahwa Ali adalah seorang maling

Pengujian Statistika yang Umum

Jenis-jenis pengujian statistika sangat banyak

1. **T-Test**
2. **ANOVA**
3. Pearson Correlation Test
4. Chi-Square
5. Wilcoxon Rank-Sum Test
6. Sign Test
7. e.t.c

T-Test

T-Test merupakan pengujian untuk rata-rata contoh menggunakan sebaran-t (t-distribution)

Biasa digunakan untuk membandingkan rata-rata 2 populasi

T-Test

Type of T-Test :

- Test for 1 Population
 $H_0: \mu = 10$ vs $H_1: \mu \neq 10$
- Test for 2 Independent Population :
 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$
- Test for 2 Paired Population
 $H_0: d = 0$ vs $H_1: d \neq 0$

T-Test

Assumption :

- Sample is normal distribution
- Or Huge Number of Sample (Central Limit Theorem)

Contoh

Question :

Kita ingin menguji apakah rata-rata amount pembelian wanita lebih dari pria? (Kriteria $\alpha = 0.05$)

Langkah 1

Hipotesa

- $H_0: \mu_{\text{wanita}} = \mu_{\text{pria}}$
- $H_1: \mu_{\text{wanita}} \neq \mu_{\text{pria}}$

Langkah 2

- Berdasarkan data sampel rata-rata amount pembelanjaan wanita sebesar 2 juta per user sedangkan untuk pria 1.8 juta per user.

Langkah 3

Dengan menggunakan t-distribution kita mendapatkan
 $p\text{-value}=0.016$

Langkah 4

Karena $p\text{-value} < \alpha$

Maka kita akan mengambil keputusan bahwa perbedaan rata-rata pembelanjaan wanita dengan pria signifikan.

Yang berarti bahwa rata-rata pembelanjaan wanita lebih besar dari pria.

ANOVA

Anova merupakan pengujian terhadap keragaman dari beberapa populasi. Pengujian menggunakan sebaran-F (F-Distribution)

Biasa digunakan untuk **membandingkan rata-rata beberapa populasi** atau untuk **feature selection** pada machine learning

ANOVA

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

H_1 = At least two means differ

ANOVA

Sederhananya

H0 : Kategori **tidak** mampu membedakan respons

H1 : Kategori mampu membedakan respons

Contoh

Question :

Kita ingin menguji apakah jumlah **click** pada **button** dengan warna biru, merah, kuning, hijau memiliki perbedaan? (Kriteria alpha = 0.05)

Langkah 1

Hipotesa

- $H_0: \mu_{\text{kuning}} = \mu_{\text{merah}} = \mu_{\text{hijau}} = \mu_{\text{biru}}$
- $H_1: \mu_i \neq \mu_j, i, j = \text{kuning, merah, biru, hijau } i \neq j$

Langkah 2

- Berdasarkan data sampel rata-rata jumlah **click** sebagai berikut.

Merah = 3000 *click*

Kuning = 2900 *click*

Hijau = 2500 *click*

Biru = 3500 *click*

Langkah 3

Dengan menggunakan uji hipotesis anova kita mendapatkan $p\text{-value}=0.07$

Langkah 4

Karena $p\text{-value} > \alpha$

Maka kita akan mengambil keputusan bahwa jumlah ***clicks*** tidak dipengaruhi oleh warna

Inferential Statistics



Hypothesis Testing

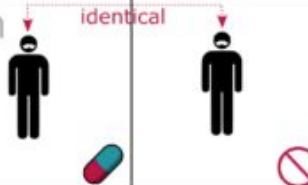


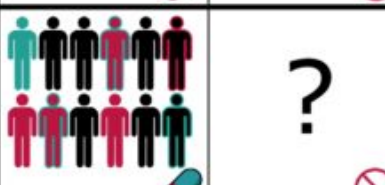


Experimental Design (A/B Testing)



Hands-On

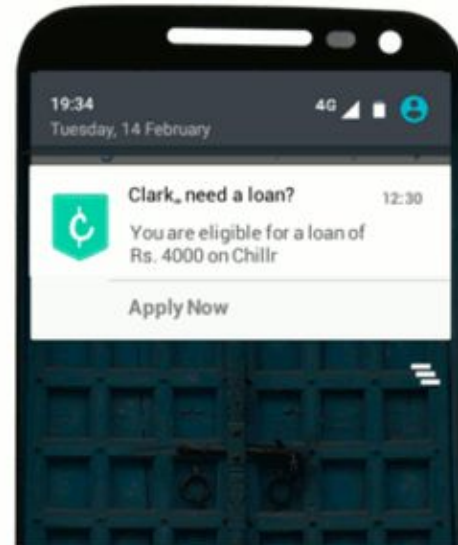
Bagaimana Mengukur Kausalitas?

Experiment		<p>Control and treatment are identical and their behavior is deterministic. Causal effect of treatment is directly the difference between observations for the two groups.</p> <p><i>Physics, Biology, Social sciences</i></p>
Statistical Experiment		<p>Control and treatment are not identical but divided at random. This makes it possible to build a precise estimate of the causal effect of treatment.</p> <p><i>A/B testing, Central Limit Theorem, Bayesian Statistics</i></p>
Quasi-experiment		<p>Control and treatment are not identical and divided by a "natural" criterion. Depending on "internal" and "external" quality of the criterion, it is possible to build a good estimate of the causal effect of treatment.</p> <p><i>Differences-in-differences, Regression Discontinuity, Instrumental variables, Matching, Controlled Regression</i></p>
Counterfactuals		<p>Control group does not exist, instead its behaviour is estimated with a predictive model of what would have happened without the treatment (= counterfactual).</p> <p><i>Synthetic Differences-in-Differences, Athey & Imbens, CausalImpact</i></p>

Stronger evidence



v/s



Steps to do an Experiment

Define an Experiment



Define Metrics



Define Duration the Experiment



Do Interim Analysis



Stop the Analysis



Post-Analysis

Define an Experiment

- **Apa nama experimentnya?**

Ex : AB Test New Design for Push Notification

- **Definisikan Hipotesis**

Ex : New design will increase the clicked rate (**click through rate**)

- **Siapa *participant*-nya?**

Ex : The user that get push notification

- **Variabel yang akan diuji apa?**

Ex : Existing Design & New Design

Define Metrics

Macroconversions

Metrics yang sangat bersentuhan dengan hipotesis

e.g. click through rate (CTR)

Microconversions

Metrics yang mengukur aksi lain yang disebabkan dari perubahan variabel

e.g. delete notification

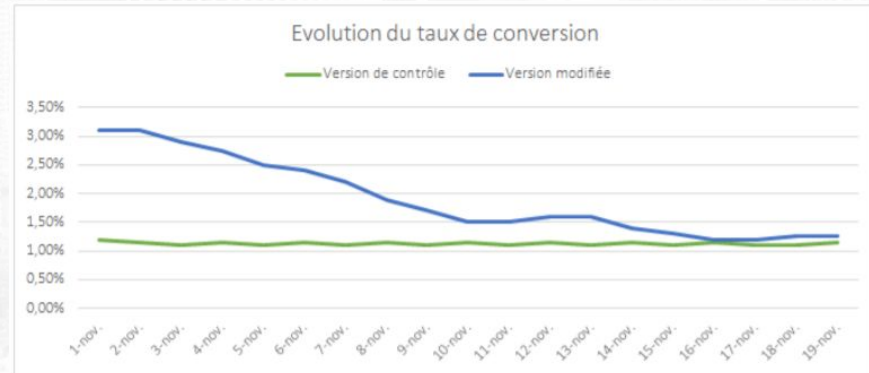
Vanity Metrics

Metrics pendukung

e.g. number of notification pushed

Define the Durations

1. Ukuran sampel
2. Seasonal Effect



Lakukan Analisa berkala

80:20



50:50



20:80



+ Jika data punya negative impact tidak hanya berdampak kepada sebagian kecil

- Keputusan tidak bisa diambil karena datanya sangat sedikit

Data sudah balance dan sudah berimbang

Mayoritas user sudah diterapkan design yang baru.

Post Analysis

Kita bisa menggunakan berbagai macam hypothesis testing (pengujian hipotesis) untuk mengambil keputusan misal

1. T-Test
2. Anova
3. Chi-Square Test
4. e.t.c

Inferential Statistics



Hypothesis Testing



Experimental Design (A/B Testing)



Hands-On



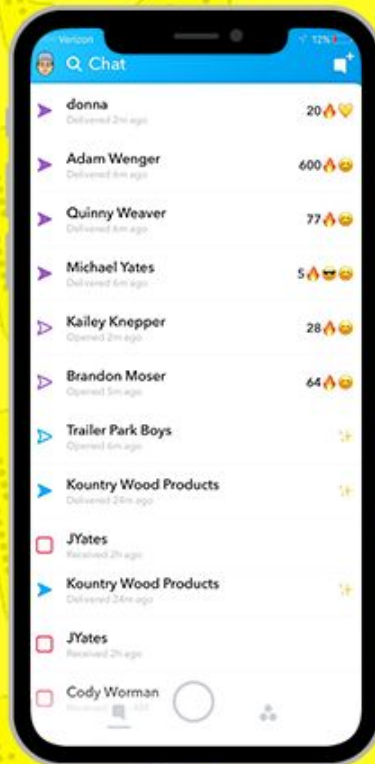
Terima Kasih!

Challenge

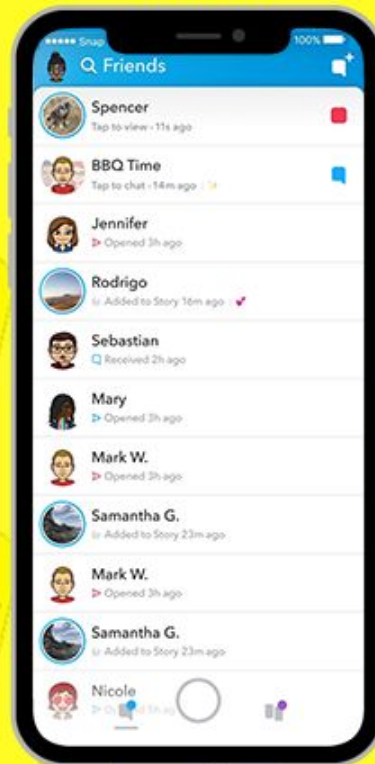
Rancang experiment untuk tampilan apps snapchat!
Dengan mengikuti step step yang telah dijelaskan.

SNAPCHAT'S OLD & NEW DESIGN OF FRIENDS SCREEN

BEFORE



AFTER



Challenge

Rancang experiment untuk tampilan apps snapchat! Silahkan teman-teman tentukan!

1. Apa nama experimentnya?
2. Definisikan Hipotesis
3. Siapa *participant*-nya?
4. Variabel yang akan diuji apa?
5. Metrics apa yang akan digunakan?
6. Berapa sample size dan durasi experiment?