twitter: @kh_notodiputro

IPB University
Bogor Indonesia

Department of Statistics
Study Program in Statistics and Data Science

Pembandingan Dua Populasi

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

email: khairil@apps.ipb.ac.id

Ketua Program Studi
Statistika dan Sains Data

Pembandingan Dua Populasi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Example 1

Making Sense of Studies Comparing Two Groups

Picture the Scenario

When e-commerce or marketing companies design webpages, there are a lot of things to consider. For instance, Amazon researched extensively which color to use for the Add To Cart button and where to place it on the site. To improve on an existing website, web designers often use so-called A/B tests in which, over a few days, some visitors to the site are routed to the existing design and others to a test site with some new design feature (such as a redesigned button) but otherwise the same functionality. Then, various metrics, such as the proportion of people who signed up for a newsletter or the amount of sales generated from each version of the site, are compared.

Questions to Explore

- How can we use data from such experiments to compare two web designs in terms of number of clicks or amount of sales generated?
- How can we use the information in the data to make an inference about the larger population of all visitors to the website? What are the assumptions we need to make for our inference to be valid?

Thinking Ahead

This chapter shows how to compare two groups on a categorical outcome (e.g., whether a visitor signed up for a newsletter) or on a quantitative outcome (e.g., the dollar amount of products a visitor puts into the shopping cart). To do this, we'll use the inferential statistical methods that the previous two chapters introduced: confidence intervals and significance tests.

For categorical variables, we will compare proportions between two groups. In Examples 2 to 4, we'll look at aspirin and placebo treatments, studying the proportions of subjects getting cancer under each treatment. Exercise 10.12 presents the result of an A/B test comparing a button that reads Sign Up to a button that reads Learn More on the 2008 fundraising website of a then relatively unknown U.S. senator, Barack Obama.

For quantitative variables, we will compare means between two groups. In Examples 6 to 8, we will examine whether including a basic graph in the description of a product leads to a higher mean rating of the product's usefulness and, in Example 9, we will investigate how the mean reaction time differs between students using a cell phone and those just listening to the radio in a simulated car-driving environment.

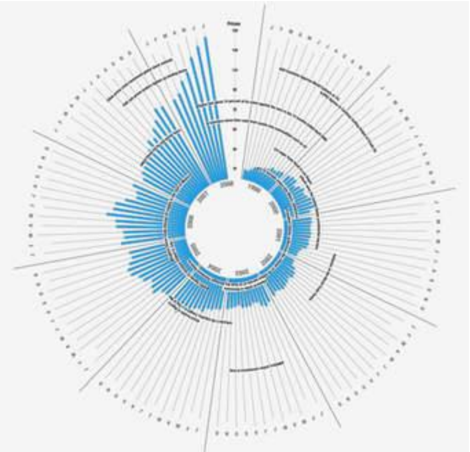
Motivasi

- **Suatu persh ingin tahu situs (*website*) spt apa yg disukai konsumen.**
- **Dibandingkan situs yg ada skr dg situs dengan rancangan baru.**
- **Diamati brp pengunjung situs lama dan situs baru, brp banyak yg login, brp lama mrk menelusuri situs, dst.**

Bisakah data pengunjung situs itu digunakan untuk penarikan kesimpulan (*inference*)? Bagaimana caranya?

Pembandingan Dua Populasi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Populasi-1 ⟷ Populasi-2

- Parameter yang akan dibandingkan? Proporsi ataukah nilaitengah?
- Apakah ragam kedua populasi sama ataukah berbeda?
- Ukuran contoh besar atau kecil?

Pembandingan Dua Populasi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Membandingkan dua populasi →

- membandingkan dua parameter proporsi → p_1 vs p_2
- membandingkan dua parameter nilaitengah → μ_1 vs μ_2

Dalam bentuk hipotesis →

- $H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$ (dua arah)
- $H_0: p_1 = p_2$ vs $H_1: p_1 > p_2$ atau $H_1: p_1 < p_2$ (satu arah)
- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$ (dua arah)
- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 > \mu_2$ atau $H_1: \mu_1 < \mu_2$ (satu arah)

Pengujian perlu asumsi/kondisi →

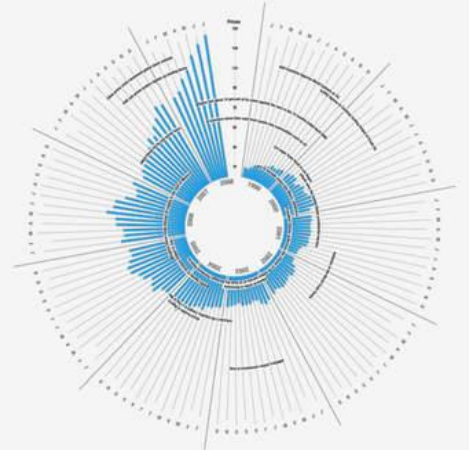
- Asumsi : kedua contoh berasal dari populasi normal
- Ragamnya sama atau tidak → $\sigma_1^2 = \sigma_2^2$ atau kah $\sigma_1^2 \neq \sigma_2^2$
- Kondisi contohnya apakah n besar atau n kecil?

→ Uji Z atau kah uji t

Catatan: Sama saja $H_0: p_1 = p_2 \Leftrightarrow H_0: p_1 - p_2 = 0$, juga $H_0: \mu_1 = \mu_2 \Leftrightarrow H_0: \mu_1 - \mu_2 = 0$

Pembandingan Dua Proporsi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Pembandingan proporsi

Populasi-1 dgn parameter p_1

p_1 vs p_2

Populasi-2 dgn parameter p_2

H_0: $p_1 = p_2$
H_1: $p_1 > p_2$
atau
H_1: $p_1 < p_2$
atau
H_1: $p_1 \neq p_2$

H_0: $p_1 - p_2 = 0$
H_1: $p_1 - p_2 > 0$
atau
H_1: $p_1 - p_2 < 0$
atau
H_1: $p_1 - p_2 \neq 0$

Statistik $\hat{p}_1$

statistik ($\hat{p}_1$ - $\hat{p}_2$)

Statistik $\hat{p}_2$

$$s_{(\hat{p}1-\hat{p}2)} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Ukuran contoh *n* besar

Statistik uji adalah $Z_{hit} = (\hat{p}_1 - \hat{p}_2) - p_0 / s_{(\hat{p}1-\hat{p}2)}$

Bandingkan Z_{hit} dengan Z_{tabel} → Tolak/terima H_0: $p_1 - p_2 = 0$

H_0 ditolak jika:

- **Satu-arah → $Z_{hit} < -Z_\alpha$ atau $Z_{hit} > Z_\alpha$**
- **Dua arah → $|Z_{hit}| > Z_{\alpha/2}$**

Pembandingan Dua Proporsi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Example 3

Cancer Death Rates for Aspirin and Placebo

Picture the Scenario

In Example 2, the sample proportions of subjects who died of cancer were $\hat{p}_1 = 347/11535 = 0.030$ for placebo (Group 1) and $\hat{p}_2 = 327/14035 = 0.023$ for aspirin (Group 2). The estimated difference was $\hat{p}_1 - \hat{p}_2 = 0.030 - 0.023 = 0.007$.

Questions to Explore

a. What is the standard error of this estimate?
b. How should we interpret this standard error?

Think It Through

a. Using the standard error formula given, we have

$$se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.030(1-0.030)}{11535} + \frac{0.023(1-0.023)}{14035}} = 0.002.$$

b. Consider all the possible experiments with 11,535 participants in the placebo group and 14,035 participants in the aspirin group, just as in the contingency table shown in Table 10.1. From each experiment, compute the difference $(\hat{p}_1 - \hat{p}_2)$ in the sample proportions between these two groups. This difference will not always equal 0.007, the difference obtained from Table 10.1, but vary from one experiment to the next. The standard error of 0.002 describes how much these differences vary around the actual (unknown) difference in the population.

Insight

From the se formula, we see that se decreases as n_1 and n_2 increase. The standard error is very small for these data because the sample sizes were so large. This means that the $(\hat{p}_1 - \hat{p}_2)$ values would be very similar from study to study. It also implies that $(\hat{p}_1 - \hat{p}_2) = 0.007$ is quite precise as an estimate of the actual difference in the population proportions.

▶ *Try Exercise 10.3, part b*

Ilustrasi pembandingan proporsi

- **Ingin diketahui efektivitas dari obat untuk menyembuhkan kanker.**
- **Sejumlah pasien kanker ada yg diberi obat *placebo* (grup 1) dan ada yg diberi obat aktif (grup 2).**
- **Statistik $\hat{p}_1$= 347/11535=0.030 dan $\hat{p}_2$=327/14035=0.023, bedanya: $(\hat{p}_1-\hat{p}_2)$=0.030-0.023=0.007**
- $s_{(\hat{p}1-\hat{p}2)} = \sqrt{\frac{0.030(1-0.030)}{11535} + \frac{0.023(1-0.023)}{14035}} = 0.002.$
- $H_0: p_1-p_2=0$ vs $H_1: p_1-p_2 \neq 0$ → Uji dua arah
- **Statistik uji Z_{hit} = (0.007-0)/0.002=3.5**
- **$Z_{0.05/2}$ = 1.96 → Tolak H_0 pada taraf nyata α = 5%. Artinya obat tsb bisa menekan kanker.** → $|Z_{hit}| > Z_{\alpha/2}$

Pembandingan Dua Proporsi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

SUMMARY: Two-Sided Significance Test for Comparing Two Population Proportions

1. Assumptions

- A categorical response variable observed in each of two groups
- Independent random samples, either from random sampling or a randomized experiment
- n_1 and n_2 are large enough that there are at least five successes and five failures in each group if using a two-sided alternative

2. Hypotheses

Null H_0: $p_1 = p_2$ (that is, $p_1 - p_2 = 0$)

Alternative H_a: $p_1 \neq p_2$ (one-sided H_a also possible; see after Example 5)

3. Test Statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} \text{ with } se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

where $\hat{p}$ is the pooled estimate.

4. P-value

P-value = Two-tail probability from standard normal distribution (Table A) of values even more extreme than observed z test statistic presuming the null hypothesis is true

5. Conclusion

Smaller P-values give stronger evidence against H_0 and supporting H_a. Interpret the P-value in context. If a decision is needed, reject H_0 if P-value ≤ significance level (such as 0.05).

Ringkasan teknis pengujian

Asumsi: data berasal dari contoh acak, n_1 dan n_2 besar.

H_0: $p_1 - p_2 = 0$ vs H_1: $p_1 - p_2 > 0$
atau H_1: $p_1 - p_2 < 0$
atau H_1: $p_1 - p_2 \neq 0$ (dua arah)

Statistik uji: Z_{hit} dengan galat baku $s_{(p1-p2)}$

Nilai-p = P(|Z| > Z_{hit}) atau P(Z > Z_{hit}) atau P(Z < Z_{hit}).

Simpulan: nilai-p lbh kecil dari α maka H_0 ditolak .

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Ilustrasi: Masalah kurang tidur

Do people who work long hours have more trouble sleeping? This question was examined in the paper "Long Working Hours and Sleep Disturbances: The Whitehall II Prospective Cohort Study" (*Sleep* [2009]: 737–745). The data in the accompanying table are from two independently selected samples of British civil service workers, all of whom were employed full-time and worked at least 35 hours per week. The authors of the paper believed that these samples were representative of full-time British civil service workers who work 35 to 40 hours per week and of British civil service workers who work more than 40 hours per week.

	n	Number who usually get less than 7 hours of sleep a night
Work over 40 hours per week	1501	750
Work 35–40 hours per week	958	407

Do these data support the theory that the proportion that usually get less than 7 hours of sleep a night is higher for those who work more than 40 hours per week than for those who work between 35 and 40 hours per week? Let's carry out a hypothesis test with $\alpha = .01$. For these samples

Misalkan

p_1 = proporsi orang yang berkerja lebih dari 40 per minggu dan tidur kurang dari 7 jam;

p_2 = proporsi orang yang berkerja 35-40 per minggu dan tidur kurang dari 7 jam;

$H_0: p_1 - p_2 = 0$ vs $H_1: p_1 - p_2 > 0$.

Taraf nyata $\alpha = 0.01$.

Statistik uji: $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

Beberapa angka penting: n_1=1501, $\hat{p}_1$= 0.500, n_2 = 958, $\hat{p}_2$= 0.425.

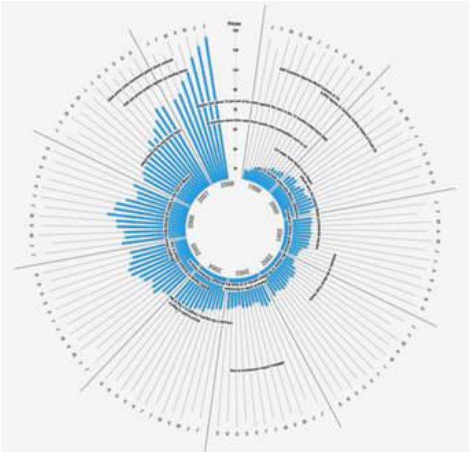
Sehingga statistik uji z = 3.64.

Padahal $z_{.01}$ = 2.33 (lihat tabel normal baku), sehingga $H_0: p_1 - p_2 = 0$ ditolak yang artinya orang yang bekerja per minggu lebih dari 40 jam biasanya tidur kurang dari 7 jam per malam.

Carilah berapa nilai-p dari pengujian ini..!!!

Pembandingan Dua Proporsi

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

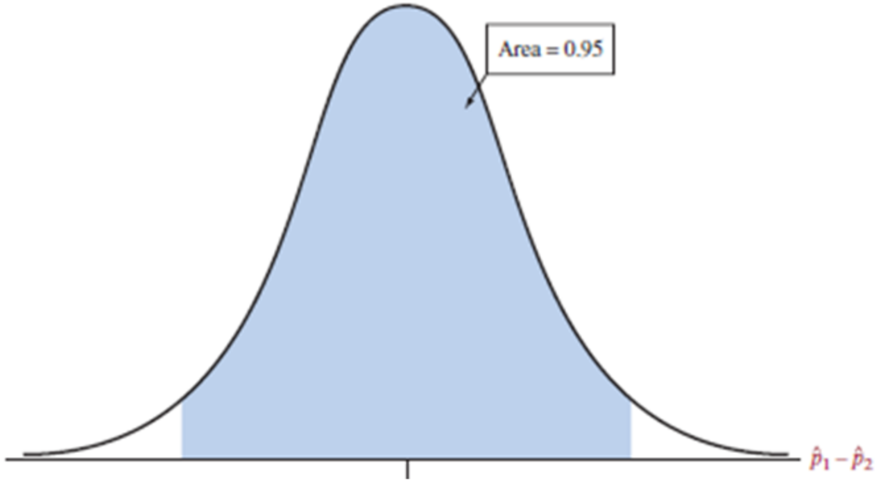


Selang kepercayaan selisih dua proporsi

Selang kepercayaan (1-α)100% bagi (p_1-p_2):

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\, s_{(\hat{p}1-\hat{p}2)}$$

dengan $s_{(\hat{p}1-\hat{p}2)} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$



Cancer death rates

Untuk kasus obat kanker maka selang kepercayaan 95% bagi (p_1-p_2):

$(\hat{p}_1 - \hat{p}_2) \pm 1.96(se)$, or $0.007 \pm 1.96(0.002)$, which is 0.007 ± 0.004, or $(0.003, 0.011)$.

Table 10.2 MINITAB Output for Confidence Interval Comparing Proportions

Number of Cancer Deaths ↓ Observed $\hat{p}$ ↓

Sample	X	N	Sample p
1	347	11535	0.030082
2	327	14035	0.023299

Difference = p(1) − p(2)
Estimate for difference: 0.00678346 ← This is $(\hat{p}_1 - \hat{p}_2)$
95% CI for difference: (0.00279030, 0.0107766)

Hasil MINITAB

Perhatikan bhw SK tersebut tidak mencakup nilai nol. Apa artinya???

Diskusi Dulu.....

Sesi 1... beres!!! → Sesi 2...

Pembandingan Dua Nilaitengah ($\sigma_1 \neq \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Pembandingan nilaitengah ($\sigma_1 \neq \sigma_2$)

Populasi-1 dgn parameter μ_1, σ_1

μ_1 vs μ_2

Populasi-2 dgn parameter μ_2, σ_2

$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 > \mu_2$	$H_1: \mu_1 - \mu_2 > 0$
atau	atau
$H_1: \mu_1 < \mu_2$	$H_1: \mu_1 - \mu_2 < 0$
atau	atau
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 - \mu_2 \neq 0$

H_0 ditolak jika:

- Satu-arah → $Z_{hit} < -Z_\alpha$ atau $Z_{hit} > Z_\alpha$
- Dua arah → $|Z_{hit}| > Z_{\alpha/2}$

Statistik $\bar{x}_1, s_1$

statistik $(\bar{x}_1 - \bar{x}_2)$

Statistik $\bar{x}_2, s_2$

$$s_{(\bar{x}1-\bar{x}2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Ragam tdk sama, n besar.

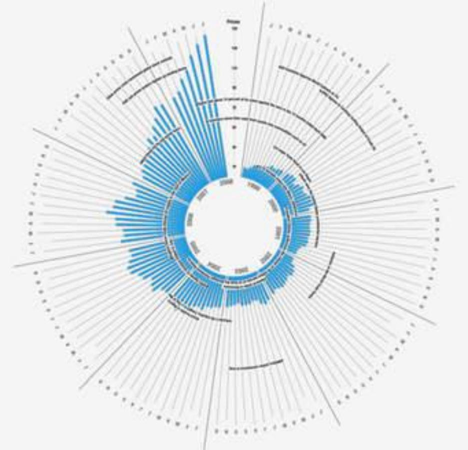
Statistik uji adalah $Z_{hit} = [(\bar{x}_1 - \bar{x}_2) - \mu_0] / s_{(\bar{x}1-\bar{x}2)}$

Bandingkan Z_{hit} vs Z_{tabel} → Tolak/terima $H_0: \mu_1 - \mu_2 = 0$

Asumsi: kedua populasi berasal dari populasi normal dengan ragam tdk sama.

Pembandingan Dua Nilaitengah ($\sigma_1 \neq \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



A Graph Is Worth a Thousand Words

Picture the Scenario

We learned in previous chapters that graphs provide a powerful way of conveying information. Public relations, media, or marketing companies all use graphs to draw attention to their services or products and ultimately boost sales. Are we giving more credence to a service or product if its description is accompanied by a graph, i.e., looks more scientific? To explore this, in a recent experiment participants were randomly split into two groups. One group just read a short generic text about the effectiveness of a drug. The other group read that same text but now accompanied by a basic bar graph that just mirrored the text and did not provide any new information.[4] After the experiment, participants rated the perceived effectiveness of the drug on a 9-point scale, with 1 representing "not at all effective" and 9 representing "very effective." Figure 10.4 shows side-by-side box plots, and Table 10.5 provides some summary statistics for the ratings in the two groups.

Table 10.5 Summary for Ratings on Perceived Effectiveness of Medication

		Ratings on perceived effectiveness	
Group	Sample Size	Mean	Standard Deviation
Text and graph	30	6.83	1.18
Text only	31	6.13	1.43

Question to Explore

How can we compare the ratings between the group that saw the text and the graph and the group that just saw the text?

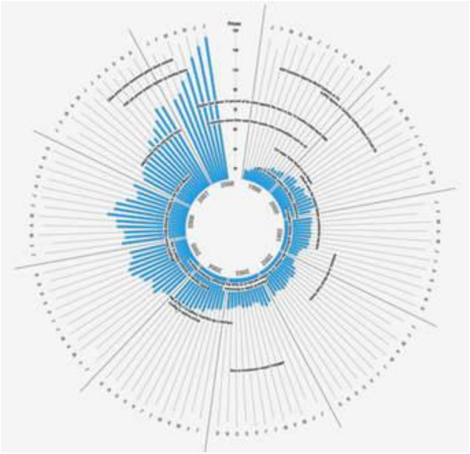
Pembandingan nilaitengah

Grafik dalam iklan

- **Ingin diketahui pengaruh grafik thd keterbacaan teks iklan obat.**
- **31 org → baca teks iklan saja**
- **30 org → baca teks iklan yg ditambahi grafik**
- **Diukur pendapat responden thd keampuhan obat tsb (skala: 1-9)**
- **Data/statistiknya spt pada Tabel 10.5.**

Pembandingan Dua Nilaitengah ($\sigma_1 \neq \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Grafik dalam iklan ($\sigma_1 \neq \sigma_2$)

Teks + grafik, parameter μ_1, σ_1

μ_1 vs μ_2

Teks saja, parameter μ_2, σ_2

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$

Hipotesis dua arah

Ragam tdk sama, *n* besar.

Statistik
$\bar{x}_1 = 6.83$
$s_1 = 1.18$

statistik ($\bar{x}_1$ - $\bar{x}_2$)

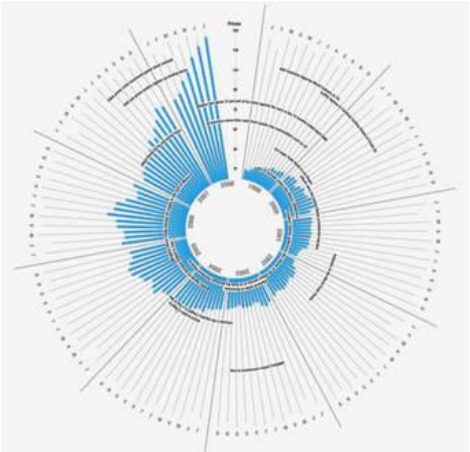
Statistik
$\bar{x}_2 = 6.13$
$s_2 = 1.43$

$$s_{(\bar{x}1-\bar{x}2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.18)^2}{30} + \frac{(1.43)^2}{31}} = 0.336$$

Statistik uji Z_{hit} = (0.7-0)/0.336 = 2.08. Ternyata $|Z_{hit}| > 1.96$ → Tolak H_0: $\mu_1 - \mu_2 = 0$ pada α=5%. Artinya grafik menambah kepercayaan konsumen thd iklan.

Pembandingan Dua Nilaitengah ($\sigma_1 \neq \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Selang Kepercayaan: Grafik dalam iklan

Teks + grafik, parameter μ_1, σ_1

μ_1 vs μ_2

Teks saja, parameter μ_2, σ_2

Ragam tdk sama, *n* besar.

```
Two-sample T for Text&Graph vs Text

              N    Mean   StDev   SE Mean
Text&Graph   30    6.83    1.18      0.21
Text         31    6.13    1.43      0.26

Difference = μ (Text&Graph) − μ (Text)
Estimate for difference: 0.704
95% CI for difference: (0.033, 1.375)
```

Statistik
$\bar{x}_1 = 6.83$
$s_1 = 1.18$

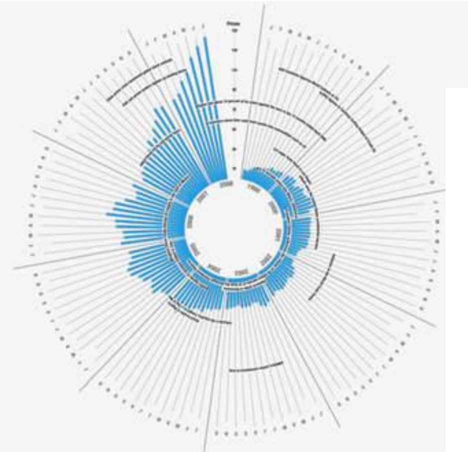
statistik ($\bar{x}_1$ - $\bar{x}_2$)

Statistik
$\bar{x}_2 = 6.13$
$s_2 = 1.43$

Selang kepercayaan 95% bagi (μ_1 - μ_2) ϵ (0.033, 1.375), selang ini tidak mencakup nilai nol shg kita 95% yakin bhw beda nilaitengah dari dua populasi itu tidak sama dengan nol. Artinya tambahan grafik di dalam iklan membuat konsumen lebih percaya adanya khasiat dari obat.

Pembandingan Dua Nilaitengah ($\sigma_1 = \sigma_2$)

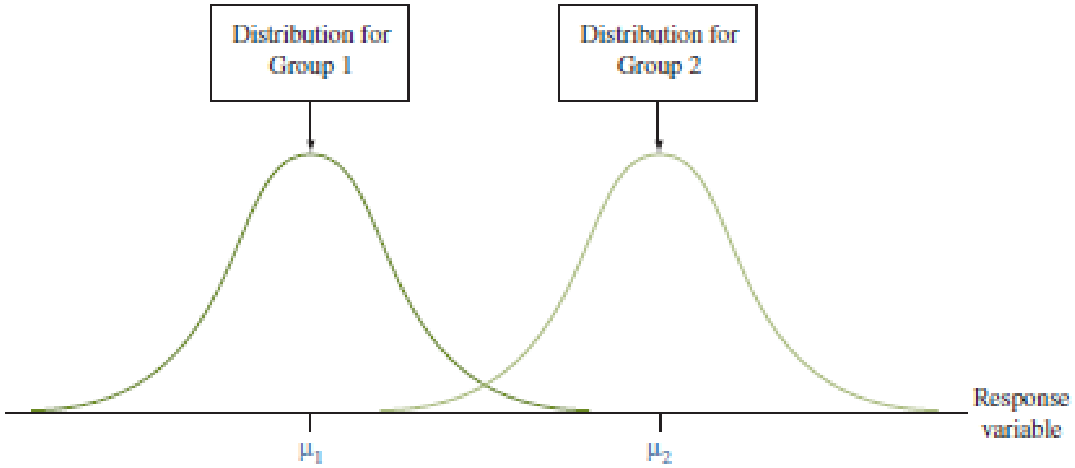
IPB University — Bogor Indonesia —
Department of Statistics
Statistics and Data Science Study Program



Comparing Means, Assuming Equal Population Standard Deviations

An alternative t method to the one described in Section 10.2 is sometimes used when, under the null hypothesis, it is reasonable to expect the *variability* as well as the mean to be the same. For example, consider a study comparing a drug to a placebo in terms of lowering blood pressure. If the drug has no effect, then we expect the entire distributions of the response variable (blood pressure) to be identical for the two groups, not just the mean. This method requires an extra assumption in addition to the usual ones of independent random samples and approximately normal population distributions:

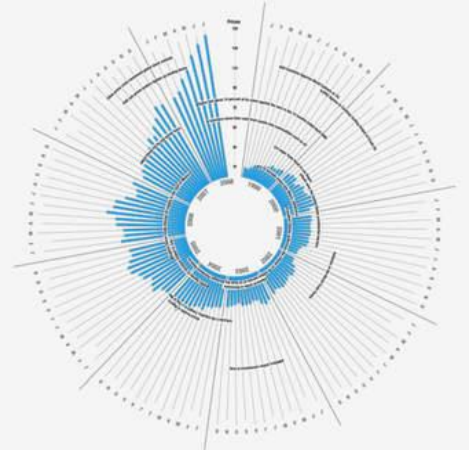
The population standard deviations are equal, that is, $\sigma_1 = \sigma_2$ (see Figure 10.7).



Asumsi: Ragam kedua populasi sama

- **Bagaimana jika ragam dua populasi yang dibandingkan dianggap sama?**
- **Pada prinsipnya sama saja, bedanya terletak pada galat baku dari statistik beda antara pop-1 dan pop-2.**
- **Galat baku pada kasus ini merupakan galat baku gabungan yang dihitung dari kedua contoh.**

Pembandingan Dua Nilaitengah ($\sigma_1 = \sigma_2$)



Pembandingan nilaitengah (σ_1=σ_2)

Populasi-1 dgn parameter μ_1,σ

μ_1 vs μ_2

Populasi-2 dgn parameter μ_2,σ

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$
atau
H_1: $\mu_1 < \mu_2$
atau
H_1: $\mu_1 \neq \mu_2$

H_0: μ_1-μ_2 =0
H_1: μ_1-μ_2 >0
atau
H_1: μ_1-μ_2 <0
atau
H_1: μ_1-μ_2 ≠0

Statistik $\bar{x}_1$, s_1

statistik ($\bar{x}_1$ - $\bar{x}_2$)

Statistik $\bar{x}_2$, s_2

Ragam sama dan *n* besar.

$$s_{(\bar{x}1-\bar{x}2)} = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \rightarrow s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

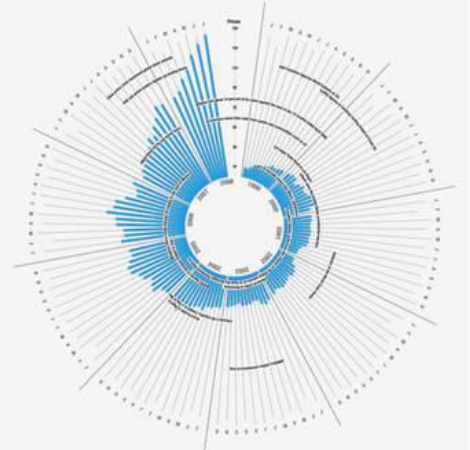
Statistik uji adalah Z_{hit} = [($\bar{x}_1$-$\bar{x}_2$)-μ_0]/$s_{(\bar{x}1-\bar{x}2)}$

Bandingkan Z_{hit} vs Z_{tabel} → Tolak H_0: μ_1 - μ_2 =0

H_0 ditolak jika:
- **Satu-arah → $Z_{hit} < -Z_\alpha$ atau $Z_{hit} > Z_\alpha$**
- **Dua arah → $|Z_{hit}| > Z_{\alpha/2}$**

Pembandingan Dua Nilaitengah ($\sigma_1 = \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



The descriptive statistics compare lavage and debridement arthroscopic surgery to a placebo (fake surgery) treatment.

Group		Knee Pain Score		
		Sample Size	Mean	Standard Deviation
Placebo	(μ_1)	60	51.6	23.7
Arthroscopic—lavage		61	53.7	23.7
Arthroscopic—debridement	(μ_2)	59	51.4	23.2

MINITAB

```
Sample      N     Mean     StDev     SE Mean
   1       60     51.6      23.7       3.1
   2       59     51.4      23.2       3.0
Difference = μ(1) − μ(2)
Estimate for difference: 0.20
95% CI for difference:   (-8.23, 8.63)
T-Test of difference = 0(vs ≠):
T-Value = 0.047  P-Value = 0.963  DF = 117
Both use Pooled StDev = 23.4535
```

Ilustrasi: Debridement vs Placebo surgery

H_0: μ_1-μ_2 =0 H_1: μ_1-μ_2 ≠0

Galat baku:

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(60-1)23.7^2 + (59-1)23.2^2}{60+59-2}} = 23.45$$

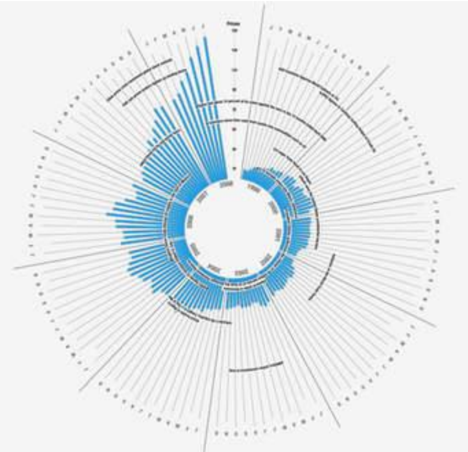
$$se = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 23.45\sqrt{\frac{1}{60}+\frac{1}{59}} = 4.30$$

Z_{hit} = (51.6-51.4)/4.30 = 0.047

Ternyata |Z_{hit} | < 1.96 → terima H_0, artinya belum cukup bukti bhw rasa nyeri sebagai hasil dari operasi *debridement* berbeda dengan *placebo*.

Pembandingan Dua Nilaitengah ($\sigma_1 = \sigma_2$)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Selang kepercayaan (1-α)100% bagi μ_1-μ_2

- Seperti apa selang kepercayaan (1-α)100% bagi $\boldsymbol{\mu_1}$-$\boldsymbol{\mu_2}$ jika ragam populasi sama?
- SK (1-α)100% bagi $\boldsymbol{\mu_1}$-$\boldsymbol{\mu_2}$ disusun mengikuti sebaran percontohan dari statistik $(\bar{X}_1 - \bar{X}_2)$ jika ukuran contoh besar dan diasumsikan ragam populasi sama.
- Lambangkan $(\bar{X}_1 - \bar{X}_2)$ sebagai D, maka D ~ Normal ($\boldsymbol{\mu_1}$-$\boldsymbol{\mu_2}$, $\boldsymbol{\sigma^2}$), sehingga SK (1-α)100% bagi $\boldsymbol{\mu_1}$-$\boldsymbol{\mu_2}$ adalah

$$d - z_{\alpha/2} s_d < \boldsymbol{\mu_1}\text{-}\boldsymbol{\mu_2} < d + z_{\alpha/2} s_d$$

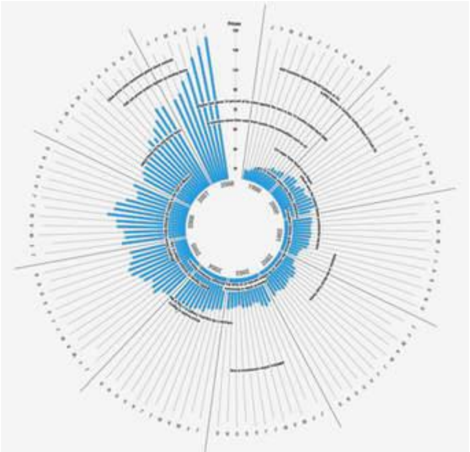
dengan $s_d = s_{(\bar{x1}\text{-}\bar{x2})}$ $= s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$. $\rightarrow$ $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$.

- Ilustrasi → SK 95% bagi selisih nilai nyeri dari operasi Debridement vs Placebo adalah: {(51.6-51.4) - (1.96)(4.30)} sampai {(51.6-51.4) + (1.96)(4.30)}.
- Hasilnya adalah ($\boldsymbol{\mu_1}$-$\boldsymbol{\mu_2}$) ∈ (-8.23, 8.63) → **Mencakup nol, artinya???**

Diskusi Dulu.....

Sesi 2... beres!!! → Sesi 3...

Pembandingan Dua Nilaitengah



Contoh Kecil

Bagaimana jika ragam dua populasi yang dibandingkan dianggap sama tetapi ukuran contohnya kecil?

Pada prinsipnya sama saja, bedanya terletak pada statistik uji yang digunakan, yaitu statistik t_{hit}.

Besarnya $t_{hit} = (\bar{x}_1 - \bar{x}_2)/s_e$ ~ t-Student dgn $db = n_1+n_2-2$.

$$s_e = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \rightarrow s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}.$$

Jika $|t_{hit}| > t_{\alpha/2(n1+n2-2)}$ maka H_0 ditolak.

$$df = \frac{(n_1-1)(n_2-1)}{(1-c)^2(n_1-1)+c^2(n_2-1)}, \quad \text{with } c = \frac{s_1^2/n_1}{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Bagaimana jika ragam dua populasi yang dibandingkan dianggap tidak sama tetapi ukuran contohnya kecil?

Pada prinsipnya sama saja, bedanya terletak pada statistik uji yang digunakan, yaitu statistik t_{hit}.

Besarnya $t_{hit} = (\bar{x}_1 - \bar{x}_2)/s_e$ ~ t-Student dgn *db* yang lebih kompleks.

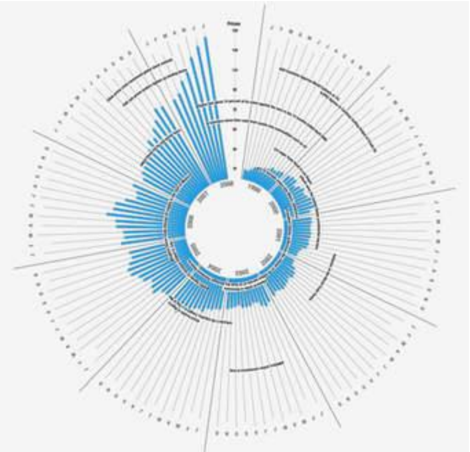
Dugaan galat baku dari beda rataan adalah

$$s_e = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Sebaran percontohannya **mendekati sebaran t** tetapi dengan derajat bebas yang agak kompleks

Pembandingan Dua Nilaitengah

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Ilustrasi: Raket Tennis Baru ($\sigma_1 \neq \sigma_2$)

There have been enormous improvements in the design of tennis rackets in the last 20 years. To investigate whether the new oversized racket delivered less stress to the elbow than a more conventionally sized racket, a group of 45 tennis players of intermediate skill volunteered to participate in the study. Because there was no current information on the oversized rackets, an unbalanced design was selected. Thirty-three players were randomly assigned to use the oversized racket and the remaining 12 players used the conventionally sized racket. The force on the elbow just after the impact of a forehand strike of a tennis ball was measured.

	Oversized	Conventional
Sample Size	33	12
Sample Mean	25.2	33.9
Sample Standard Deviation	8.6	17.4

Ragam dua kali lipat

$$H_0:\ \mu_1 - \mu_2 \geq 0 \quad \text{versus} \quad H_a:\ \mu_1 - \mu_2 < 0$$

$$\text{T.S.:}\quad t' = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(25.2 - 33.9) - 0}{\sqrt{\dfrac{(8.6)^2}{33} + \dfrac{(17.4)^2}{12}}} = -1.66$$

$$c = \frac{s_1^2/n_1}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} = \frac{(8.6)^2/33}{\dfrac{(8.6)^2}{33} + \dfrac{(17.4)^2}{12}} = .0816$$

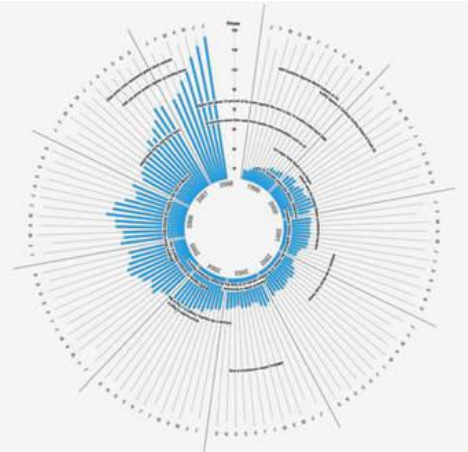
$$\text{df} = \frac{(n_1 - 1)(n_2 - 1)}{(1 - c^2)(n_1 - 1) + c^2(n_2 - 1)}$$

$$= \frac{(33 - 1)(12 - 1)}{(1 - .0816)^2(33 - 1) + (.0816)^2(12 - 1)} = 13.01$$

Untuk α = 0.05, nilai $t_{0.05(13)}$= 1.771 artinya kita menolak H_0 jika t_{hit} < -1,771. Faktanya nilai t_{hit} = -1.66 sehingga H_0 diterima. Hasil pengujian ini tidak nyata, artinya tidak cukup bukti untuk mengatakan raket baru ini bisa mengurangi tekanan pada pundak. → INKONKLUSIF

Contoh Berpasangan (*related samples*)

Department of Statistics
Statistics and Data Science Study Program



Contoh takbebas

Selama ini kita menganggap bhw contoh yang kita miliki berasal dari dua populasi yang bebas. Dalam praktek adakalanya contoh yg kita peroleh terkait satu sama lain. Bagaimana menganalisis data seperti ini?

d_i

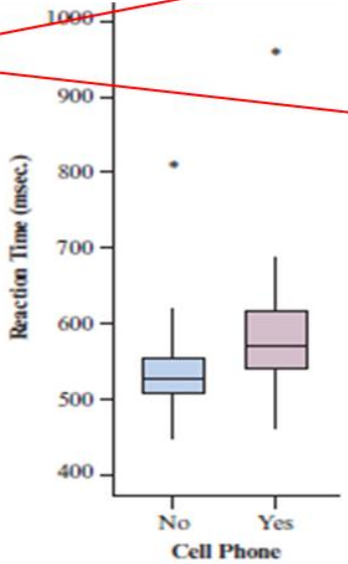


Table 10.16 shows the mean of the reaction times (in milliseconds) for each subject under each condition. Figure 10.11 in the margin shows box plots of the data for the two conditions.

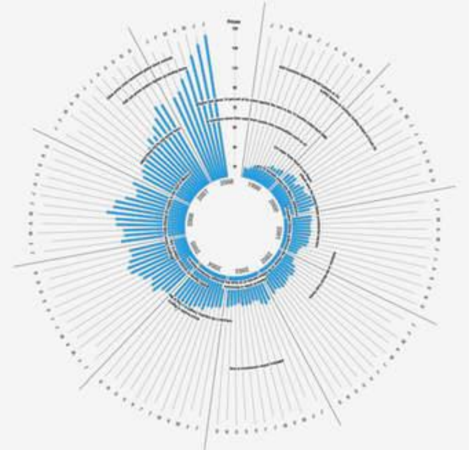
Table 10.16 Reaction Times on Driving Skills Before and While Using Cell Phone

The difference score is the reaction time using the cell phone minus the reaction time not using it, such as 636 − 604 = 32 milliseconds.

	Using Cell Phone?				Using Cell Phone?		
Student	No	Yes	Difference	Student	No	Yes	Difference
1	604	636	32	17	525	626	101
2	556	623	67	18	508	501	−7
3	540	615	75	19	529	574	45
4	522	672	150	20	470	468	−2
5	459	601	142	21	512	578	66
6	544	600	56	22	487	560	73
7	513	542	29	23	515	525	10
8	470	554	84	24	499	647	148
9	556	543	−13	25	448	456	8
10	531	520	−11	26	558	688	130
11	599	609	10	27	589	679	90
12	537	559	22	28	814	960	146
13	619	595	−24	29	519	558	39
14	536	565	29	30	462	482	20
15	554	573	19	31	521	527	6
16	467	554	87	32	543	536	−7

Contoh Berpasangan (*related samples*)

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



Contoh takbebas

Hipotesis:

H_0: $\mu_d = 0$ vs H_1: $\mu_d \neq 0$

Statistik uji: $t = \dfrac{\bar{x}_d - 0}{se}$

$$se = s_d/\sqrt{n}.$$

$$s_d = \sqrt{[(d_1-\bar{x}_d)^2+\ldots+(d_n-\bar{x}_d)^2]/(n-1)}$$

Paired T for Cell phone − Pre-cell phone

	N	Mean	StDev	SE Mean
Cell	32	585.2	89.6	15.8
Pre-cell	32	534.6	66.4	11.7
Difference	32	50.63	52.49	9.28

95% CI for mean difference: (31.70, 69.55)

T-Test of mean difference = 0 (vs ≠ 0):

T-Value = 5.46 P-Value = 0.000

Compare Means with Matched Pairs: *Use Paired Differences*

d = reaction time using cell phone − reaction time without cell phone.

For Subject 1, $d = 636 - 604 = 32$. Table 10.16 also shows these 32 difference scores. The sample mean of these difference scores, denoted by $\bar{x}_d$, is

$$\bar{x}_d = (32 + 67 + 75 + \cdots -7)/32 = 50.6.$$

The sample mean difference is $\bar{x}_d = 50.6$, and the standard deviation of the difference scores is $s_d = 52.5$. The standard error is $se = s_d/\sqrt{n} = 52.5/\sqrt{32} = 9.28$. The t test statistic for the significance test of H_0: $\mu_d = 0$ (hence equal population means for the two conditions) against H_a: $\mu_d \neq 0$ is

$$t = \bar{x}_d/se = 50.6/9.28 = 5.46.$$

With 32 difference scores, $df = n - 1 = 31$. Table 10.17 reports the two-sided P-value of 0.000. There is extremely strong evidence that the population mean reaction times are different.

For a 95% confidence interval for $\mu_d = \mu_1 - \mu_2$, with $df = 31$, $t_{.025} = 2.040$. We can use $se = 9.28$ from part a. The confidence interval equals

$$\bar{x}_d \pm t_{.025}(se), \text{ or } 50.6 \pm 2.040(9.28),$$

which equals 50.6 ± 18.9, or $(31.7, 69.5)$.

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

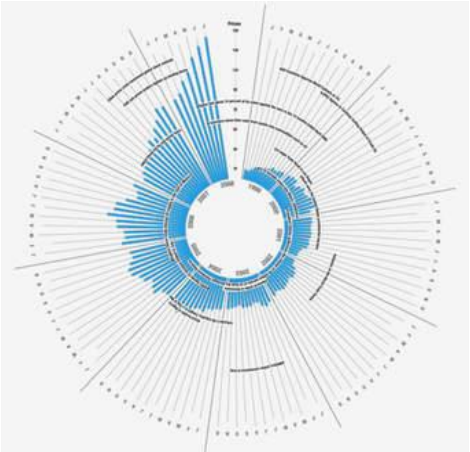
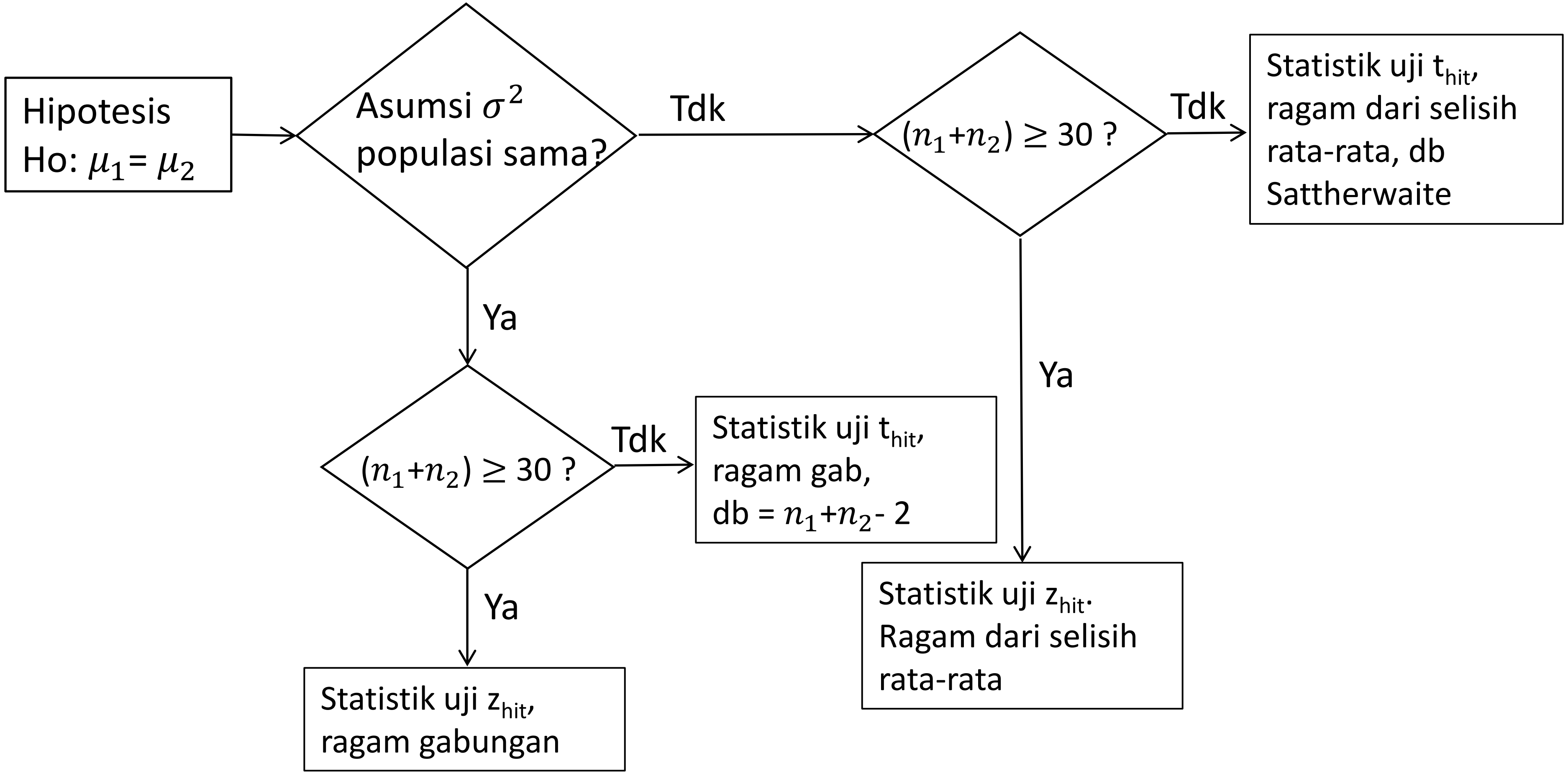


Diagram alir pengujian hipotesis untuk dua populasi



Penutup Pengujian Hipotesis

IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program

Pengujian Hipotesis

Satu Populasi

- **Ingin menguji satu parameter populasi, misal μ saja atau σ^2 saja**
- **Pengujian bisa menggunakan uji Z (Normal) atau uji t-Student dengan db = *n-1***
- **Penggunaan uji Z atau t-Student tergantung diketahui atau tidaknya σ^2 atau tergantung besar atau kecilnya ukuran contoh *n***

Dua Populasi

- **Ingin membandingkan parameter dari dua populasi, misal membandingkan μ_1 dan μ_2.**
- **Pengujian bisa menggunakan uji Z (Normal) atau uji t-Student dengan db = $n_1 + n_2 - 2$ atau db Satterthwaite**
- **Penggunaan uji Z atau t-Student tergantung sama atau tidaknya ragam kedua populasi atau tergantung besar atau kecilnya ukuran contoh *n***

db = derajat bebas atau degree of freedom

☺ THANK YOU ☺

Department of Statistics

Study Programs in Statistics and Data Science