

Biostatistics

Ferdian Fadly

2024-03-12

Table of contents

Kata Pengantar	4
1 Pendahuluan	5
1.1 Notasi-notasi/symbol yang digunakan dalam modul ini	5
1.2 Perbedaan Proporsi vs Persentase vs Rasio vs <i>Rate</i>	5
1.2.1 Proporsi	5
1.2.2 Persentase	5
1.2.3 Rasio	6
1.2.4 Rate	6
1.3 Nilai Tengah	6
1.3.1 Mean (\bar{X})	6
1.3.2 Median (Me)	7
1.3.3 Perbandingan Mean dan Median	7
1.4 Tipe-tipe Variabel	7
1.4.1 Variabel <i>Outcome</i> vs Variabel <i>Exposure</i>	7
1.4.2 Variabel <i>Outcome</i>	8
1.4.3 Variabel <i>Exposure</i>	8
2 Estimasi	9
2.1 Parameter	9
2.1.1 Tipe-tipe Parameter	9
2.1.2 Tujuan	9
2.2 Penyajian Hasil Estimasi	9
2.2.1 Estimasi Titik	9
2.2.2 Estimasi Interval atau <i>Confidence Interval</i>	9
2.3 Contoh Estimasi Selang Kepercayaan	10
2.4 Tipe-tipe Studi/Penelitian	10
2.4.1 <i>Cross-Sectional Study</i>	10
2.4.2 <i>Cohort Study</i>	11
2.4.3 <i>Case-Control Study</i>	11
2.4.4 <i>Randomised Controlled Trial (RCT)</i>	11
3 Binary Variable	12
3.1 Estimasi Proporsi π	12

3.2	Membandingkan 2 Proporsi	13
3.2.1	Contingency Table	13
3.2.2	Contoh Kasus	13
3.2.3	Risk Difference (RD)	14
3.2.4	Relative Risk (RR)	14
3.2.5	Odds dan Odds Ratio	14
3.2.6	Odds Ratio vs Risk Ratio	15
4	Summary	16
	References	17

Kata Pengantar

Website ini untuk membahas Mata Kuliah Biostatistics.

1 Pendahuluan

1.1 Notasi-notasi/symbol yang digunakan dalam modul ini

- μ : mu merepresentasikan rata-rata dari sebuah distribusi data
- σ : sigma merepresentasikan standard deviasi
- λ : lambda merepresentasikan *rate*
- α : alpha sering digunakan dalam persamaan
- β : beta sering digunakan dalam persamaan

1.2 Perbedaan Proporsi vs Persentase vs Rasio vs *Rate*

1.2.1 Proporsi

Sebagian (*share*) dari keseluruhan.

e.g.

Pada sebuah penelitian terhadap 200 orang, 15 diantaranya flu. **Proporsi** yang terkena flu adalah

$$\frac{15}{200} = \frac{3}{40} = 0.075$$

1.2.2 Persentase

Proporsi yang diekspresikan dengan 100.

e.g.

Pada sebuah penelitian terhadap 200 orang, 15 diantaranya flu. **Persentase** yang terkena flu adalah

$$\frac{15}{200} \times 100 = 0.075 \times 100 = 7.5\%$$

1.2.3 Rasio

Perbandingan dari 2 buah angka. Biasanya, dipisahkan oleh “:” atau *colon*.

e.g.

Pada sebuah penelitian terhadap 200 orang, 15 diantaranya flu. **Rasio** yang terkena dan tidak terkena flu adalah

$$15 : 185 = 3 : 37$$

1.2.4 Rate

Rate menyatakan banyaknya kejadian dalam rentang waktu tertentu.

e.g.

I run 5 kilometers in half an hour. I run at the **rate** of 10 km/h.

Incidence Rate menyatakan banyaknya kejadian dalam sebuah populasi dalam rentang waktu tertentu.

e.g.

In a study of 1000 people, 12 develop cancer over one year of the study. **The Incidence Rate** is 12/1000 people/year.

1.3 Nilai Tengah

1.3.1 Mean (\bar{X})

$$\text{Mean} = \bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$$

e.g.

Dalam sebuah penelitian terhadap 10 orang, usia mereka adalah sebagai berikut :

56, 64, 5, 63, 67, 59, 66, 69, 62, 57

```
x=c(56,64,5,63,67,59,66,69,62,57)
mean(x)
```

[1] 56.8

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{56 + 64 + 5 + 63 + 67 + 59 + 66 + 69 + 62 + 57}{10} = 56.8$$

1.3.2 Median (Me)

Nilai tengah dari sebuah kumpulan angka ketika diurutkan dari yang terkecil ke yang terbesar. Ketika banyaknya angka adalah genap, maka diambil rata-rata dari kedua angka yang berada di tengah-tengah data.

Misalnya: Dalam sebuah penelitian terhadap 10 orang, usia mereka adalah sebagai berikut :

56, 64, 5, 63, 67, 59, 66, 69, 62, 57

Setelah diurutkan menjadi:

```
x=c(56,64,5,63,67,59,66,69,62,57)
sort(x)
```

[1] 5 56 57 59 62 63 64 66 67 69

Sehingga Mediannya menjadi $\frac{62+63}{2} = 62.5$

1.3.3 Perbandingan Mean dan Median

- Mean = 56.8
- Median = 62.5
- adanya nilai *outlier* akan menggeser nilai **Mean** tapi tidak demikian halnya dengan **Median**

1.4 Tipe-tipe Variabel

1.4.1 Variabel *Outcome* vs Variabel *Exposure*

Table 1.1: Nama istilah lainnya

Variabel <i>Outcome</i>	Variabel <i>Exposure</i>
Variabel Respons	<i>Risk Factor</i>
Variabel Terikat (<i>Dependent</i>)	Variabel Bebas (<i>Independent</i>)
	Variabel Penjelas (<i>Explanatory</i>)

Seringkali sebuah penelitian bertujuan untuk mengetahui hubungan antar variabel *Outcome* dan variabel *exposure*-nya

1.4.2 Variabel *Outcome*

Variabel ini dapat berupa:

- Kualitatif e.g. status kesehatan: sehat/tidak
- Kuantitatif e.g. Tekanan darah

1.4.3 Variabel *Exposure*

Variabel yang memengaruhi Variabel *Outcome*. Variabel ini dapat berupa:

- Kualitatif e.g. Menjadi Wanita yang bekerja: Ya/Tidak
- Kuantitatif e.g. Usia

2 Estimasi

2.1 Parameter

Ukuran dari suatu **Populasi**

2.1.1 Tipe-tipe Parameter

- Population Mean
- Population Proportion
- Probability
- Rate
- Odds Ratio

2.1.2 Tujuan

Utamanya tujuan sebuah penelitian atau *study* adalah untuk *estimasi* karakteristik populasi dengan menghitung karakteristik dari sampel terkait.

2.2 Penyajian Hasil Estimasi

2.2.1 Estimasi Titik

Misalnya untuk estimasi Population Mean (μ), kita menggunakan sample mean (\bar{X}).

2.2.2 Estimasi Interval atau *Confidence Interval*

$$\begin{aligned} \bar{X} &\sim \text{Normal}(\mu, \text{se}(\bar{X})) \\ Pr(\mu - 1.96 \times \text{se}(\bar{X}) < \bar{X} < \mu + 1.96 \times \text{se}(\bar{X})) &= 0.95 \end{aligned} \tag{2.1}$$

Oleh karena itu, Interpretasi yang benarnya adalah

“We are 95% confident that the CI contains the true μ value, because it was constructed in a way so that in repeated sampling 95% of CIs do contain μ ”

2.3 Contoh Estimasi Selang Kepercayaan

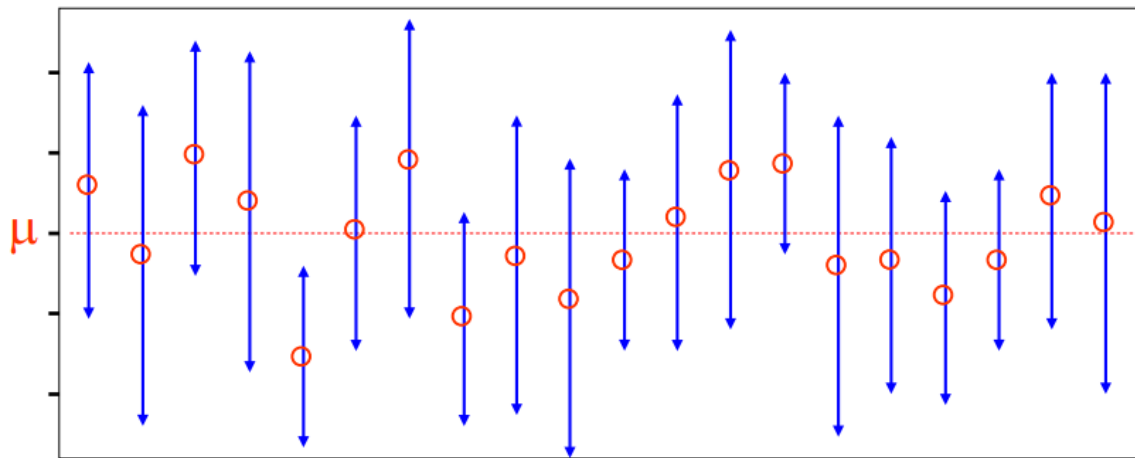
20 berat janin yang berusia 40 minggu memiliki statistik sebagai berikut:

- sample mean: $\bar{X} = 3159$ grams
- sample standard deviation: $s = 220$ grams
- So, standard errornya adalah

$$se = \frac{s}{\sqrt{n}} = \frac{220}{\sqrt{20}} = 49.2$$

95% *Confidence Interval* nya adalah:

$$(3159 - 1.96 \times 49.2, 3159 + 1.96 \times 49.2) = (3063, 3255)$$



2.4 Tipe-tipe Studi/Penelitian

- *Cross-Sectional Study* - *Cohort Study* - *Case-Control Study* - *Randomised Controlled Trial (RCT)*

2.4.1 *Cross-Sectional Study*

Pengumpulan *outcome and exposure variables* pada sebuah sampel dilakukan saat bersamaan.

2.4.1.1 Advantages

- Relatif Murah
- Dapat Mencapai tingkat partisipasi yang tinggi
- Memperoleh hasil dengan cepat

2.4.1.2 Disadvantages

- Terbatas dengan apa yang akan diestimasi
- dapat memperoleh hubungan (*association*) tapi bukan hubungan sebab-akibat (*causal*)

2.4.2 Cohort Study

Longitudinal studies yang diamati beberapa periode

2.4.2.1 Advantages

- Mengatasi masalah dari *Cross-sectional Study*

2.4.2.2 Disadvantages

- Biaya sangat mahal
- Waktu yang lama

2.4.3 Case-Control Study

Pada bentuk penelitian ini, individu penelitian diambil sebagai contoh berdasarkan status dari *outcome variable* nya

- biasa digunakan pada penyakit dengan tingkat kejadian yang rendah
- Mengambil sample individual tanpa penyakit sebagai **control**
- Variabel outcome nya bersifat *binary (case/control)*

2.4.4 Randomised Controlled Trial (RCT)

Study dimana sampel penelitian dialokasikan secara random mana yang akan memperoleh intervensi mana yang tidak (*placebo*)

(Kirkwood and Sterne 2010)

3 Binary Variable

Variabel yang nilainya cuma ada 2 (sukses/gagal, ya/tidak, sehat/tidak).

With binary outcomes we can estimate the proportion who have an outcome (disease)

i.e. we can estimate the probability or risk that an individual in a given population will have/develop this outcome

Examples:

- What proportion of Australian adults have ischaemic heart disease?

3.1 Estimasi Proporsi π

sample proportion (p)

$$p = \frac{d}{n}$$

Contoh:

- Sample size (n): 11,247
- Number with diabetes (d): 844
- Proportion

$$\begin{aligned} p &= d/n \\ &= 844/11,247 \\ &= 0.075 \end{aligned} \tag{3.1}$$

Percentage People with diabetes = 7.5% (or 75 per 1000 people).

3.2 Membandingkan 2 Proporsi

With binary outcomes we can compare estimates of the proportion with disease in different populations

i.e. we can examine if there is an association between an exposure and disease by comparing the proportions with disease in the exposed and the unexposed groups.

Examples:

- Are men more likely than women to have diabetes?

3.2.1 Contingency Table

<i>Exposure</i>	<i>Outcome (Yes)</i>	<i>Outcome (No)</i>	<i>Total</i>
<i>Yes</i>	d_1	h_1	n_1
<i>No</i>	d_2	h_2	n_2
<i>Total</i>	d	h	n

3.2.2 Contoh Kasus

You survey 330 people at the festival.

- 250 ate ice cream and 80 didn't
- Of those that ate ice cream, 170 developed gastro and 80 didn't
- Of those that didn't eat ice cream, 10 developed gastro and 70 didn't

<i>Eat Ice Cream</i>	<i>Gastro (Yes)</i>	<i>Gastro (No)</i>	<i>Total</i>
<i>Yes</i>	170	80	250
<i>No</i>	10	70	80
<i>Total</i>	180	150	330

Overall risk of gastroenteritis

$$180/330 = 0.545(54.5\%)$$

Risk of gastroenteritis in those who ate ice cream =

$$170/250 = .68(68\%)$$

Risk of gastroenteritis in those who did not eat ice cream =

$$10/80 = .125(12.5\%)$$

3.2.3 Risk Difference (RD)

$$\text{Risk Difference (RD)} = \text{Risk in exposed} - \text{Risk in unexposed}$$

3.2.4 Relative Risk (RR)

$$\text{Relative Risk (RR)} = \frac{\text{Risk in exposed}}{\text{Risk in unexposed}}$$

These are known as measures of effect size.

$$RD = 68\% - 12.5\% = 55.5\%$$

i.e. The absolute difference in risk of GI disease between those who did and did not eat the ice-cream was 55.5 percentage points.

$$RR = \frac{68\%}{12.5\%} = 5.4$$

i.e. Those who ate ice cream were 5.4 times more likely to develop GI symptoms than those who did not eat the ice cream.

3.2.5 Odds dan Odds Ratio

The odds of an event are the probability that it does happen, divided by the probability that it does not happen.

We used odds when dealing with the rare outcome.

<i>Eat Ice Cream</i>	<i>Gastro (Yes)</i>	<i>Gastro (No)</i>	<i>Total</i>	Odds of Gastro
Yes	170	80	250	170/80 = 2.125
No	10	70	80	10/70 = 0.143
Total	180	150	330	

Thus,

$$\text{Odds Ratio} = \frac{\text{Odds (Exposure=1)}}{\text{Odds (Exposure=0)}} = \frac{2.125}{0.143} = 14.9$$

Interpretation: The odds of gastroenteritis was nearly 15 times higher in those that ate ice cream compared to those that didn't

3.2.6 Odds Ratio vs Risk Ratio

Risk Ratios are usually easier to interpret and understand...

BUT Odds Ratios have some nice properties:

- we can use them for case-control studies (where there are usually not enough data to calculate risks)
- identical conclusions if we look at occurrence or absence of an outcome (e.g. odds of case versus odds of control)
- don't run into computational difficulties with common outcomes
- for a rare outcome, the Odds Ratio is very close to the Risk Ratio

4 Summary

In summary, this book has no content whatsoever.

`1 + 1`

[1] 2

References

Kirkwood, B. R., and J. A. C. Sterne. 2010. *Essential Medical Statistics*. Essentials. Wiley.
<https://books.google.co.id/books?id=8GemDQAAQBAJ>.