IPB University
Bogor Indonesia

Department of Statistics
Study Program in Statistics and Data Science

Pendugaan Parameter dan Selang Kepercayaan

twitter: @kh_notodiputro

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

email: khairil@apps.ipb.ac.id

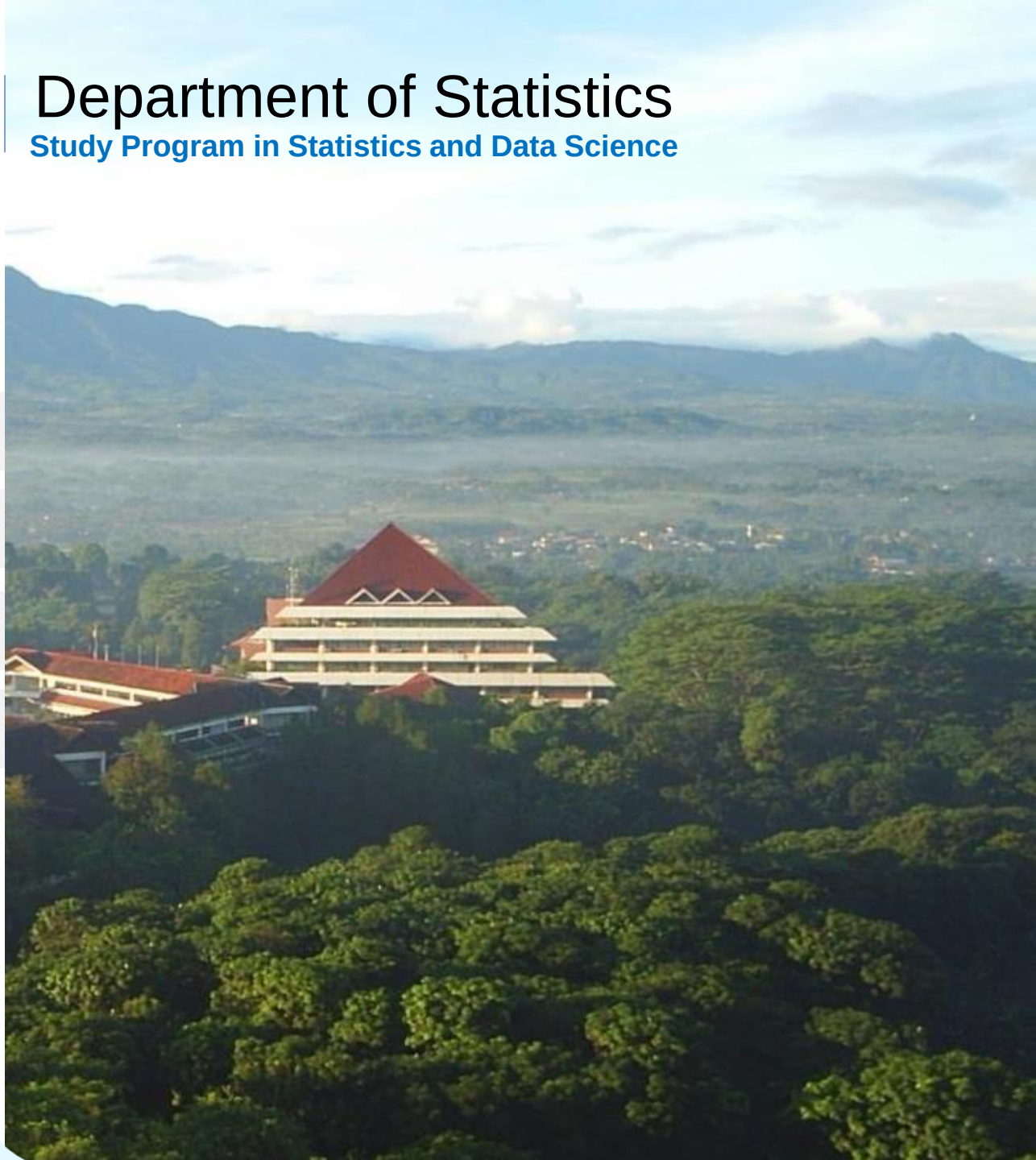
Ketua Program Studi
Statistika dan Sains Data

IPB University
Bogor Indonesia

Department of Statistics
Study Program in Statistics and Data Science

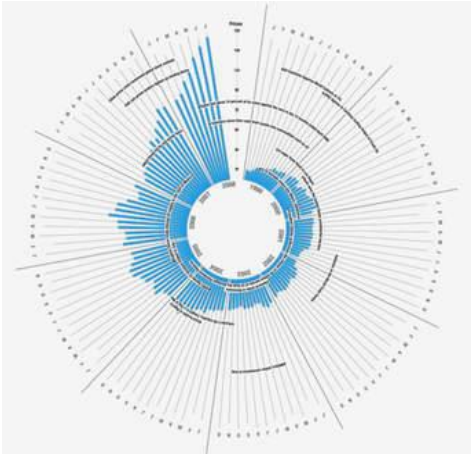
Outline

1. Kenormalan Data
2. Pendugaan Parameter
3. Selang Kepercayaan
4. Penentuan Ukuran Contoh

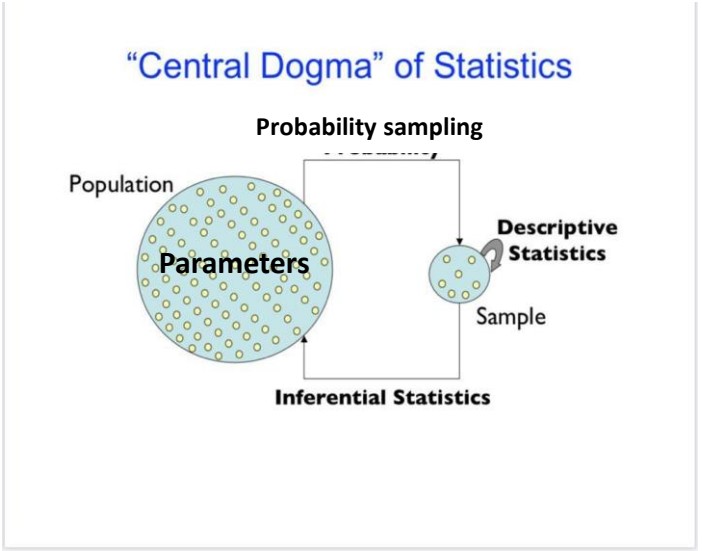


Pendugaan Parameter

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

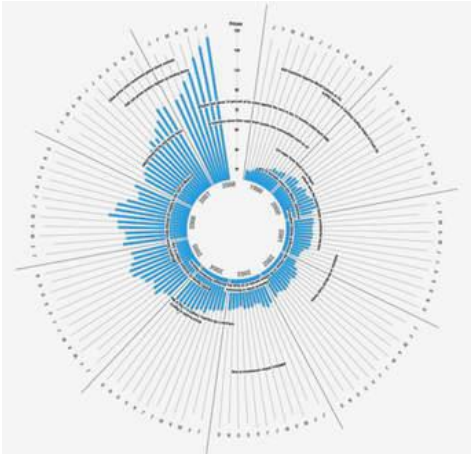


- Parameter merupakan karakteristik dari populasi
- Besaran atau angka yang ada di dalam populasi disebut parameter → statistik ada di dalam contoh
- Besarnya parameter pada umumnya tdk diketahui
- Bagaimana cara mengetahui populasi?
- Ingat ibunda ketika memasak gulai. Beliau ingin mengetahui rasa gurih dari gulai yang sedang dimasak di dalam panci. Apa yang dilakukan ibunda untuk mengetahui rasa gurih? Dicicipi…!!!
- Ilustrasi ini sangat relevan dengan pendugaan parameter populasi.→ Ibunda menduga parameter populasi berdasarkan contoh gulai yang disendoknya.



Pendugaan Parameter

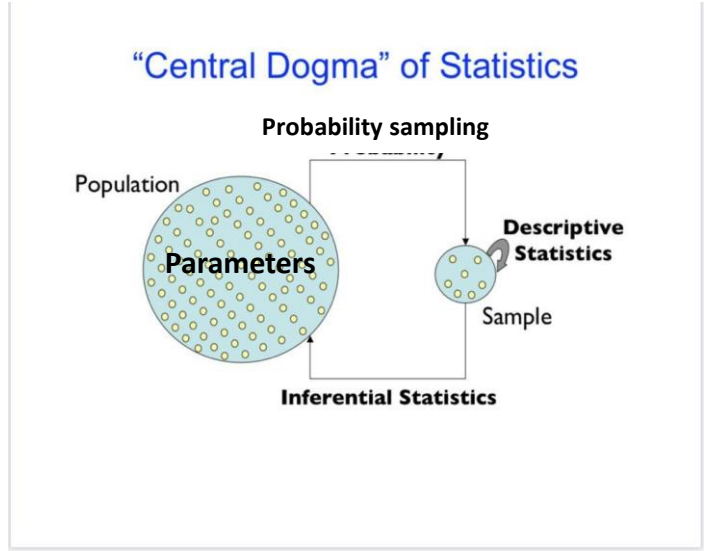
IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



- Besarnya parameter tetap tetapi besarnya statistik bisa berubah-ubah tergantung contoh yg terambil
- Satu angka penduga parameter disebut penduga titik
- Penduga parameter adalah statistik yang diperoleh dari contoh

DEFINITION

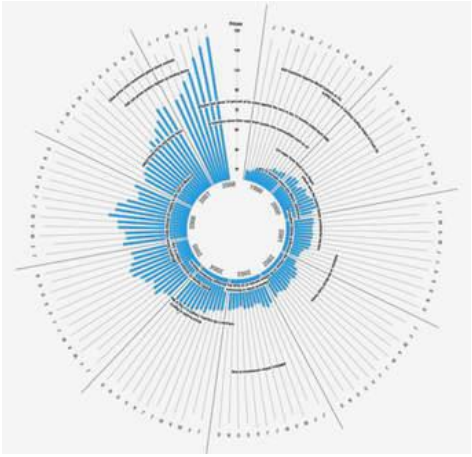
A point estimate of a population characteristic is a single number that is based on sample data and represents a plausible value of the characteristic.



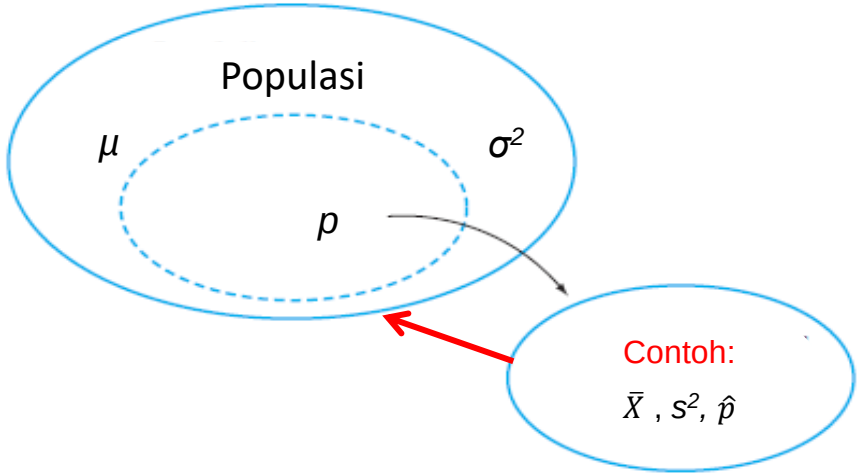
- Parameter μ diduga oleh statistik $\bar{X}$
- Parameter σ^2 diduga oleh statistik s^2
- Statistik yang terbaik adalah yang bersifat linear, takbias, dan ragam minimum → BLUE

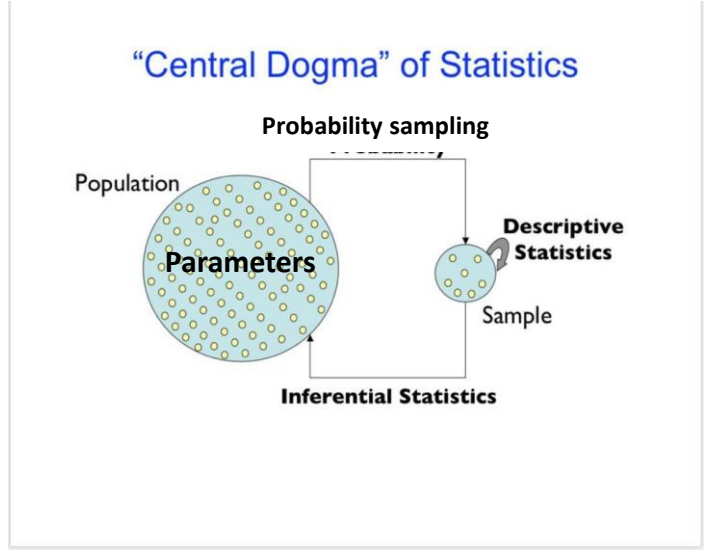
Pendugaan Parameter

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



- Perhatikan posisi parameter dan statistik:

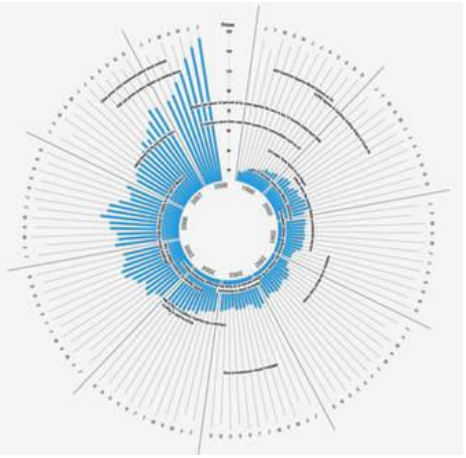




- Berapa persen mhs yang menggunakan internet lebih dari 3 jam per hari? Hasil survei menunjukkan bhw 7421 mhs ternyata ada 2998 yang berinternet > 3 jam.
- Parameter p diduga dengan statistik $\hat{p}$ → $\hat{p} = \frac{2998}{7421} = .404$
- JADI → μ, σ^2 dan p adalah PARAMETER, sedangkan $\bar{X}$, s^2, dan $\hat{p}$ adalah STATISTIK.

Pendugaan Parameter

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

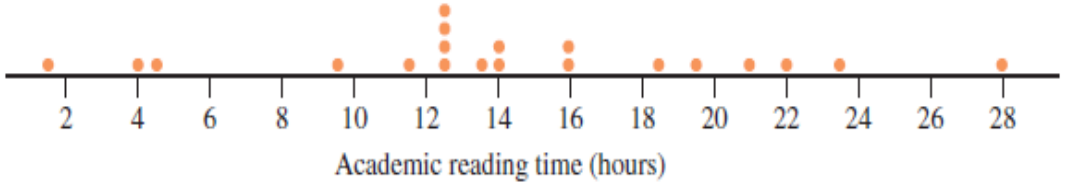


- *Academic reading*:

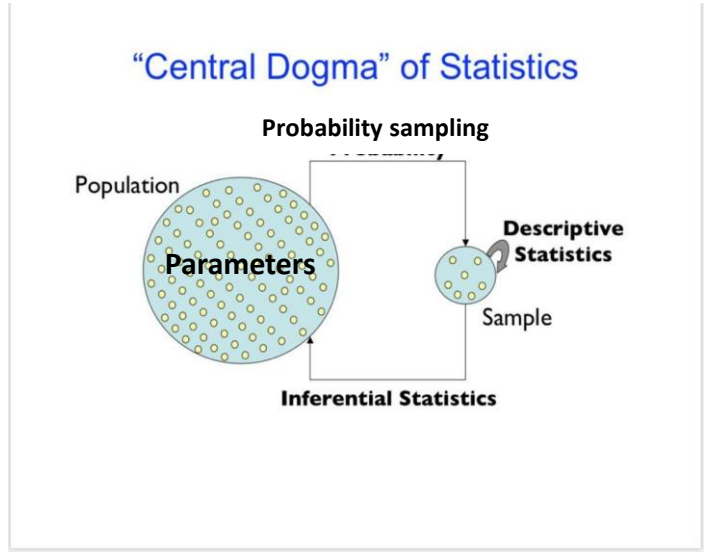
"The Impact of Internet and Television Use on the Reading Habits and Practices of College Students" (*Journal of Adolescent and Adult Literacy* [2009]: 609–619)

- Data jumlah jam penggunaan internet/minggu dari 20 mhs:

1.7	3.8	4.7	9.6	11.7	12.3	12.3	12.4	12.6	13.4
14.1	14.2	15.8	15.9	18.7	19.4	21.2	21.9	23.3	28.2

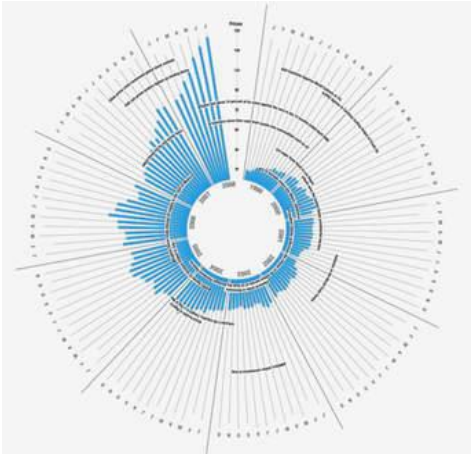


Perhatikan plot di atas → sebaran data hampir simetrik



Pendugaan Parameter

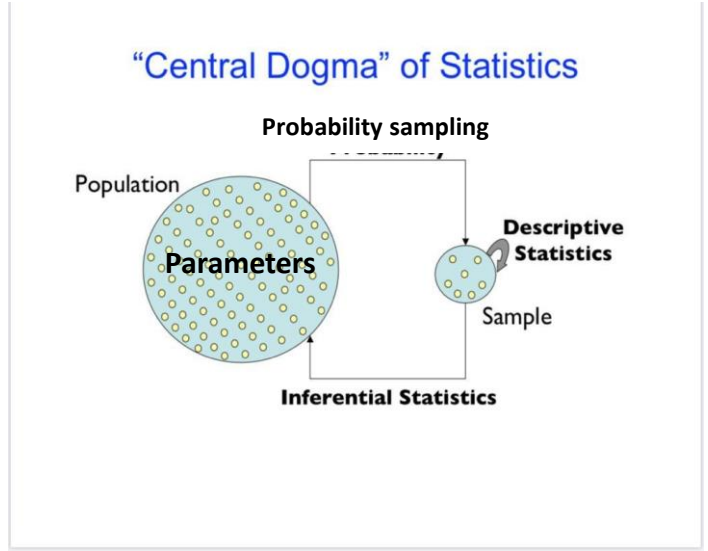
IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



- Dari plot titik → sebaran data hampir simetrik
- Parameter μ dapat diduga oleh statistik:
  - Rataan → $\bar{x} = \frac{\sum x}{n} = \frac{287.2}{20} = 14.36$
  - Median → $\text{median} = \frac{13.4 + 14.1}{2} = 13.75$

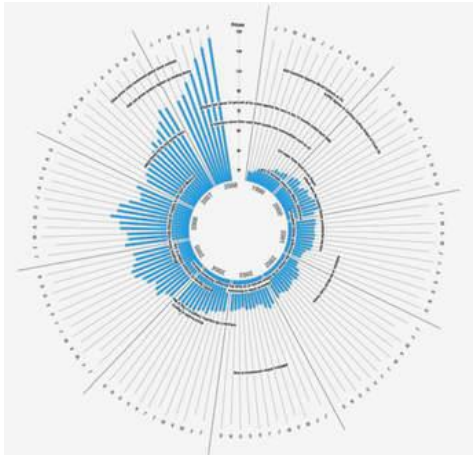
Rataan dan median hampir sama.

- Pilih statistik yang mana untuk menduga parameter? Pilih yang BLUE jika memungkinkan
- BLUE = *Best Linear Unbiased Estimator* → (*best*: ragam minimum, *linear*: mudah, *unbiased*: takbias)

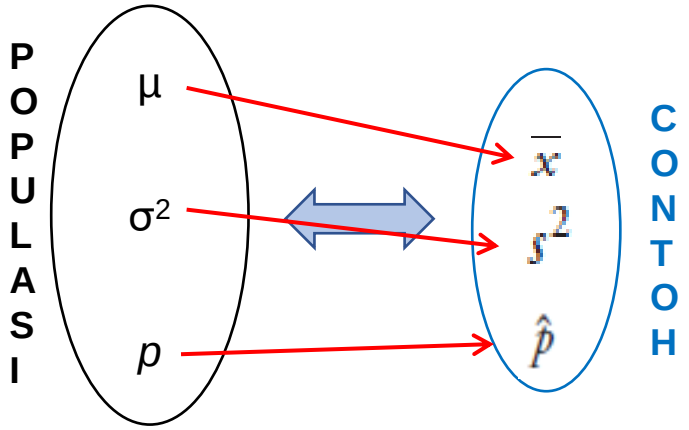


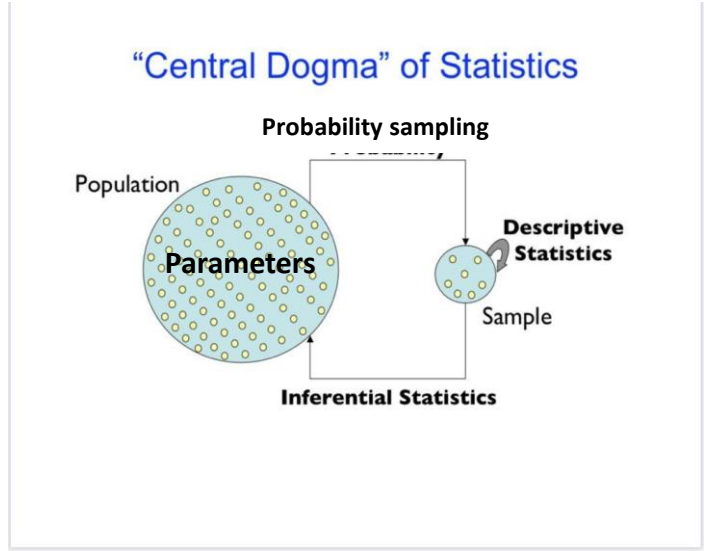
Pendugaan Parameter

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



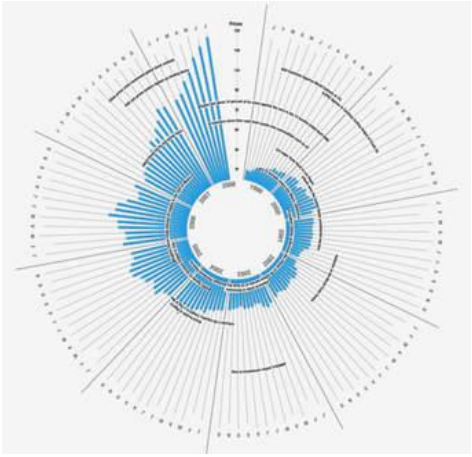
- Berikut ini penduga yg bersifat BLUE:
- Penduga bagi μ $\leftrightarrow$ $\bar{x} = \frac{\sum x}{n}$
- Penduga bagi σ^2 $\leftrightarrow$ $s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$
- Penduga bagi p $\leftrightarrow$ $\hat{p} = \frac{\text{number of successes in the sample}}{n}$



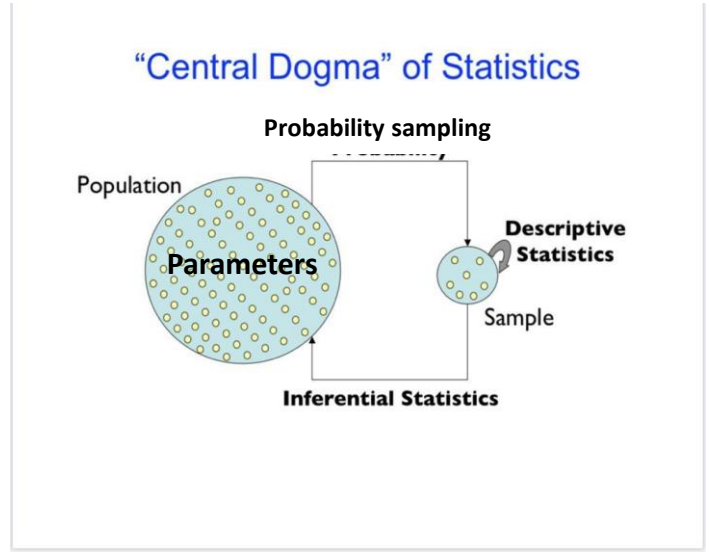


Pendugaan Parameter

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



- Penduga takbias??
  - Istilah takbias (*unbiased*) merupakan konsep jangka panjang atau berulang-ulang (*long run*), bukan konsep sesaat atau sekali 'tembak' (*short run*);
  - Jadi jika $\bar{X}$ takbias bagi μ maka tidak berarti bhw $\bar{X} = \mu$;
  - Yang benar adalah: jika $\bar{X}$ takbias bagi μ maka $E(\bar{X}) = \mu$
- Benarkah $\bar{x}$ merupakan penduga takbias bagi μ?
  - $E(\bar{X}) = E(\frac{X_1 + \ldots + X_n}{n})$
  - $E(\bar{X}) = \frac{1}{n} E(X_1 + \ldots + X_n)$
  - $E(\bar{X}) = \frac{1}{n} \{E(X_1) + \ldots + E(X_n)\}$
  - $E(\bar{X}) = \frac{1}{n} \{\mu + \ldots + \mu\} = \frac{1}{n} \{n\,\mu\} = \mu$
- Jadi $E(\bar{X}) = \mu \rightarrow \bar{x}$ takbias terhadap μ; ingatlah bhw $\bar{x}$ adalah statistik sedangkan μ adalah parameter



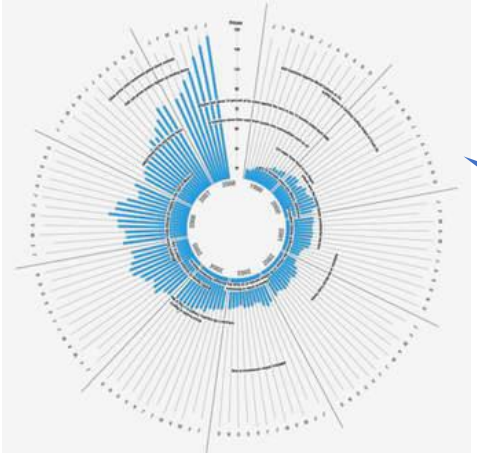
Diskusi Dulu.....

Sesi 1… beres!!!

Sesi 2…

Penduga Selang (Selang Kepercayaan)

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

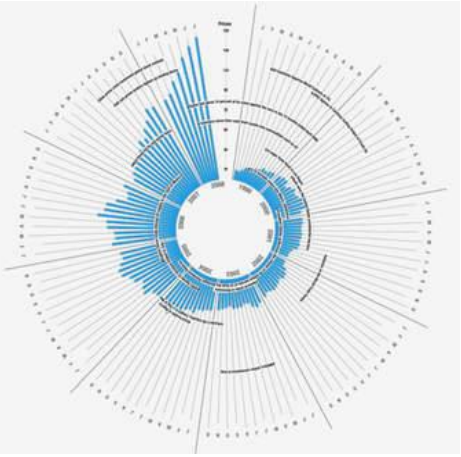


Penduga parameter atau statistik yang dihasilkan ternyata tidak memuaskan karena statistik itu pada umumnya tidak sama dengan parameternya.

→ Gunakan selang kepercayaan atau penduga selang.

Pendugaan Parameter

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



- INGAT → μ, σ^2 dan p adalah PARAMETER, sedangkan $\bar{X}$, s^2, dan $\hat{p}$ adalah STATISTIK.
- $\bar{X}$, s^2, dan $\hat{p}$ disebut sebagai penduga titik (*point estimates*) bagi parameter μ, σ^2 dan p.

Sample	Value of $\bar{y}$	Sample	Value of $\bar{y}$
2, 3	2.5	3, 10	6.5
2, 4	3	3, 11	7
2, 5	3.5	4, 5	4.5
2, 6	4	4, 6	5
2, 7	4.5	4, 7	5.5
2, 8	5	4, 8	6
2, 9	5.5	4, 9	6.5
2, 10	6	4, 10	7
2, 11	6.5	4, 11	7.5
3, 4	3.5	5, 6	5.5
3, 5	4	5, 7	6
3, 6	4.5	5, 8	6.5
3, 7	5	5, 9	7
3, 8	5.5	5, 10	7.5
3, 9	6	5, 11	8

$(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)$

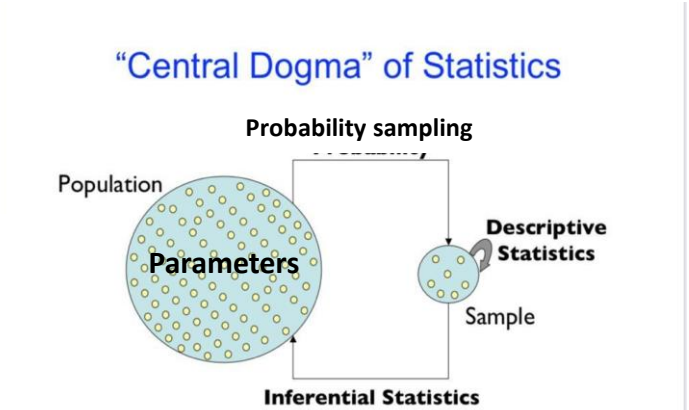
Parameter
$\mu = 6.5$; $\sigma^2 = 8.25$

Hipotetik

Contoh
n=2 → 45
kombinasi

Sample	Value of $\bar{y}$	Sample	Value of $\bar{y}$
6, 7	6.5		
6, 8	7		
6, 9	7.5	7, 11	9
6, 10	8	8, 9	8.5
6, 11	8.5	8, 10	9
7, 8	7.5	8, 11	9.5
7, 9	8	9, 10	9.5
7, 10	8.5	9, 11	10
		10, 11	10.5

45 kemungkinan



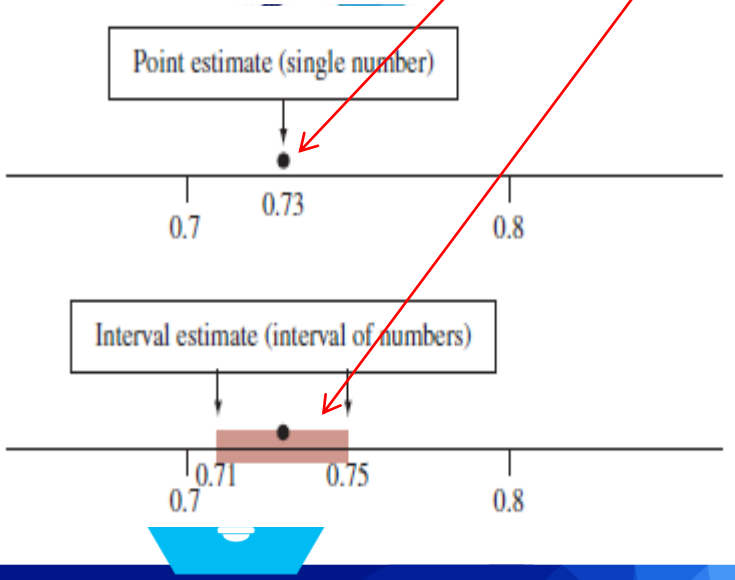
Ada berapa statistik $\bar{y}$ yang sama dengan μ ???

Selang Kepercayaan

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

A **sampling distribution** specifies the possible values a statistic can take and their probabilities. ◀

We use the **proportion** to summarize the *relative frequency* of observations in a category for a categorical variable. The proportion equals the number in the category divided by the sample size. We use the **mean** as one way to summarize the *center* of the observations for a quantitative variable. ◀



Ada dua jenis penduga bagi suatu parameter :

- **Penduga titik: suatu nilai yang merupakan tebakan terbaik terhadap parameter populasi.**
- **Penduga selang: suatu selang yang diyakini (dengan tingkat keyakinan kurang dari 100%) mencakup nilai parameter.**

Penduga titik kurang bermanfaat → tidak ada informasi seberapa dekat penduga itu dengan parameternya.

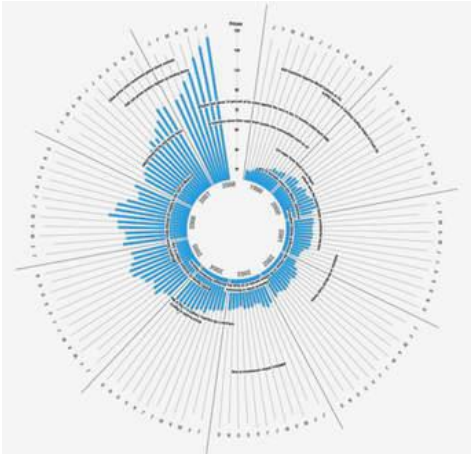
Penduga selang lebih bermanfaat → ada informasi dugaan proporsi 0.73 dapat bergeser dari parameter *p* sebesar *MoE* yaitu ±0.02.

Misal dari suatu survei terhadap mahasiswa IPB tentang pendapatnya terhadap program "*green campus*" diperoleh proporsi setuju sebesar 0.73:

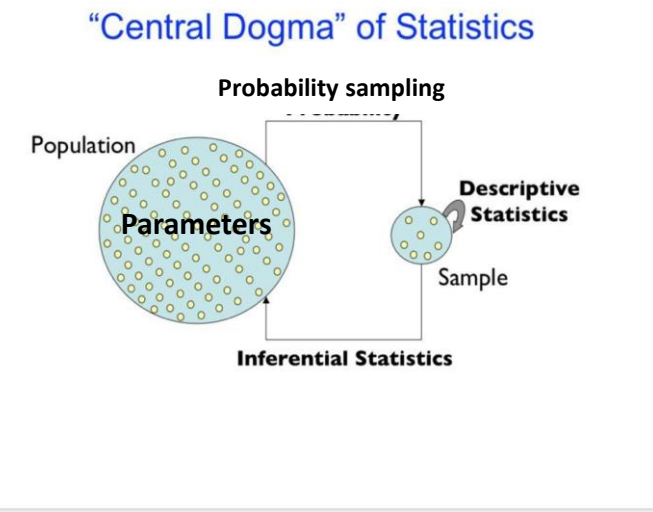
- **Penduga titik bagi parameter *p* adalah 0.73.**
- **Penduga selangnya bisa diperoleh sebesar (0.71, 0.75)**

Selang Kepercayaan

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

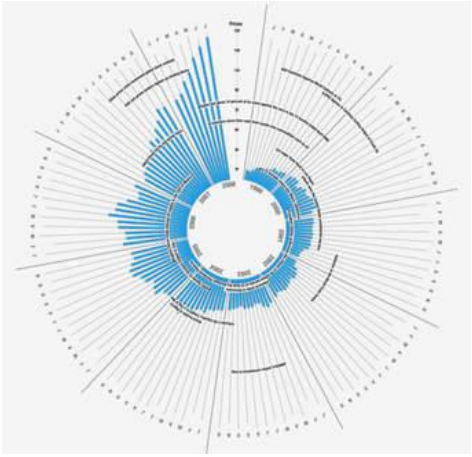


- Misalkan ada survei terhadap mahasiswa STK 2013-2017 → rataan IP TPB yaitu 3.40.
- Apakah hasil ini masuk akal? Terlalu rendah atau terlalu tinggi?
- Rataan adalah penduga titik bagi nilaitengah parameter IP TPB seluruh mahasiswa STK.
- Setiap penduga titik mungkin dekat atau mungkin jauh dengan parameter yang sesungguhnya. → perlu tahu seberapa jauh/dekat?
- Penduga selang menunjukkan ketelitian (*precision*) melalui selang yang melingkupi dugaan titik.
- Semakin sempit selang tsb semakin tinggi ketelitiannya.



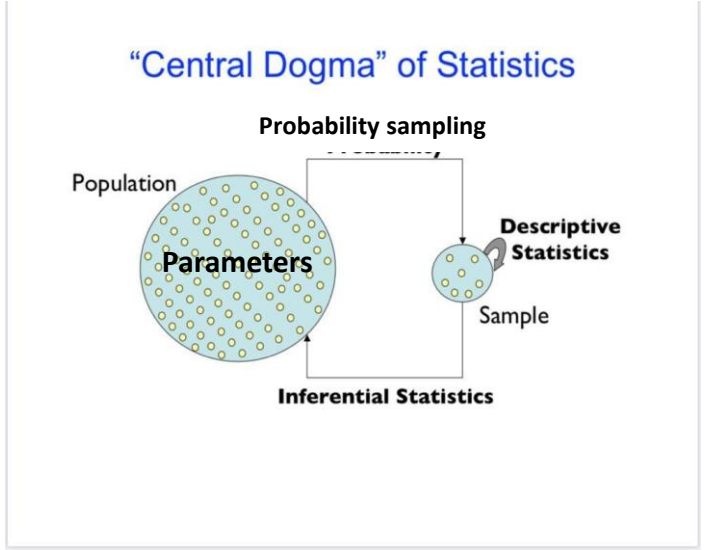
Selang Kepercayaan

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



Confidence Interval

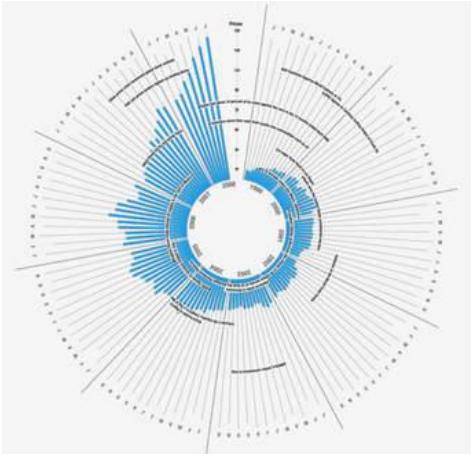
A confidence interval is an interval containing the most believable values for a parameter. It is formed by a method that combines a point estimate with a margin of error. The probability that this method produces an interval that contains the parameter is called the confidence level. This is a number chosen to be close to 1, most commonly 0.95.



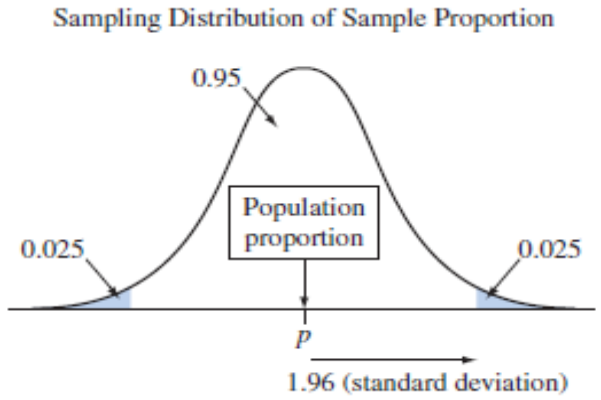
- Selang kepercayaan adalah selang nilai yang dipercaya mencakup nilai parameter yang sesungguhnya. Selang ini merupakan kombinasi dari penduga titik dengan MoE.
- Tingkat kepercayaan bagi selang ini haruslah tinggi (mendekati 1), biasanya ditentukan sebesar 0.95 atau 0.99.

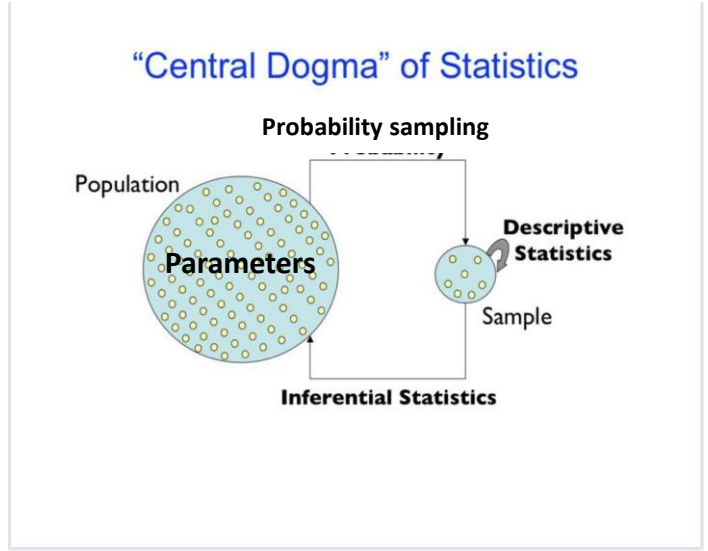
Selang Kepercayaan

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



- Selang kepercayaan (SK) dibentuk berdasarkan sebaran dari statistik (*sampling distribution*).
- Mari kita buat SK dari parameter proporsi (*p*).

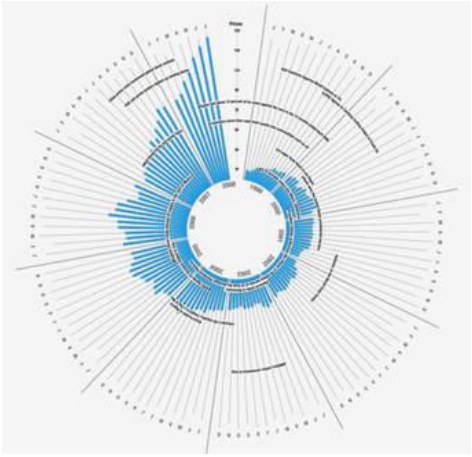


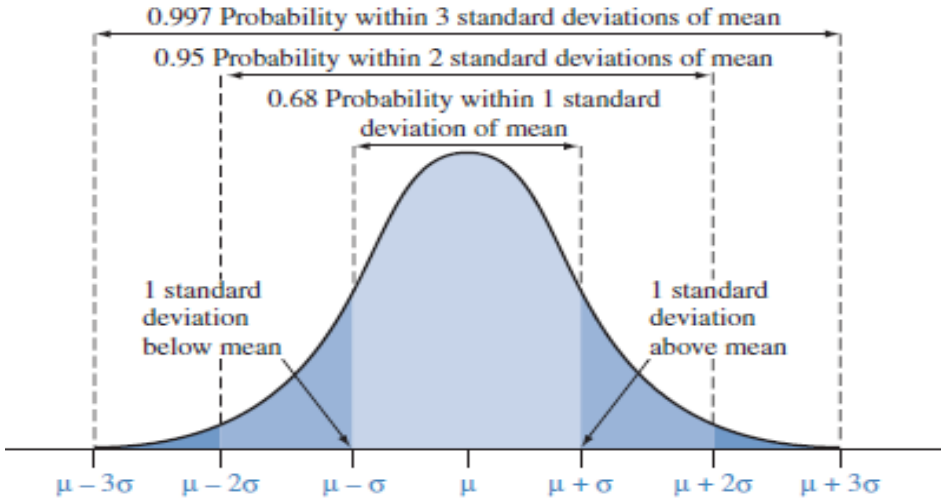


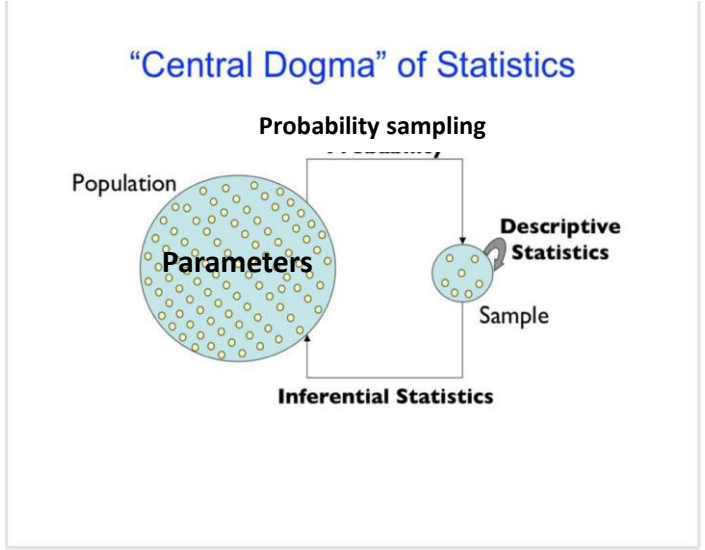
- Perhatikan bhw sebaran statistik $\hat{p}$ (dari contoh):
- Menunjukkan berbagai kemungkinan nilai $\hat{p}$;
- Sebaran $\hat{p}$ mendekati normal jika ukuran contoh *n* besar;
- Nilaitengah dari $\hat{p}$ adalah *p*;
- Simpangan baku dari $\hat{p}$ adalah $\sqrt{\frac{p(1-p)}{n}}$

Selang Kepercayaan

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





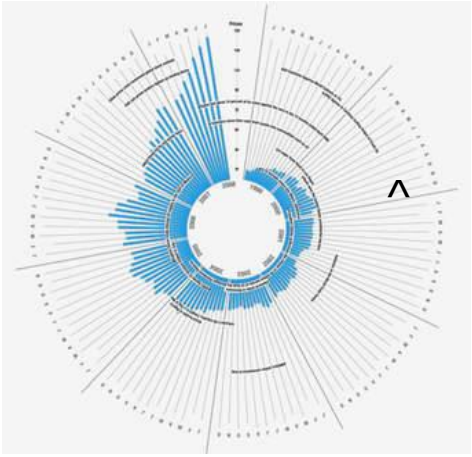


Karena sebaran dari statistik $\hat{p}$ mendekati normal:

- peluang $\hat{p}$ terletak antara -1.96σ sampai +1.96σ sama dengan 0.95 (lihat tabel normal);
- besaran 1.96σ dikenal sebagai MoE untuk selang kepercayaan 0.95 atau 95%.

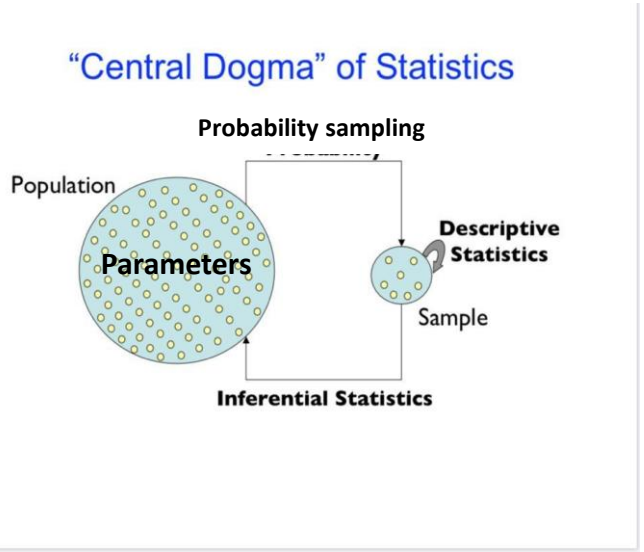
Selang Kepercayaan

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



- Jadi secara terori, peluang parameter *p* terletak antara $(\hat{p} - 1.96\sigma_{\hat{p}})$ sampai $(\hat{p} + 1.96\sigma_{\hat{p}})$ sebesar 0.95.
- Dengan kata lain, peluang selang $(\hat{p} \pm 1.96\sigma_{\hat{p}})$ mencakup parameter *p* adalah sebesar 95%.
- Selang $(\hat{p} \pm 1.96\sigma_{\hat{p}})$ disebut selang kepercayaan 95% bagi *p*, dengan $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- Dengan demikian jika dari data kita memperoleh dugaan dari *p*, yaitu $\hat{p}$ maka selang kepercayaan 95% untuk *p* dapat disusun menggunakan rumus:

$$\hat{p} \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$



Dalam praktek *p* tidak diketahui shg simpangan baku dari $\hat{p}$ dihitung dari contoh, yaitu $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Selang Kepercayaan

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

Example 2

A Wife's Career

Picture the Scenario

One question on the General Social Survey asks whether you agree or disagree with the following statement: "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." In the 2012 GSS, 31% of 1285 respondents agreed. So the sample proportion agreeing was 0.31. From a formula in the next section, we'll see that this point estimate has an estimated standard deviation of 0.01.

Questions to Explore

a. Find and interpret the margin of error for a 95% confidence interval for the population proportion who agreed with the statement.
b. Construct the 95% confidence interval and interpret it in context.

Think It Through

a. The margin of error for a 95% confidence interval for a population proportion equals 1.96 × (standard deviation), or 1.96(0.01), approximately 0.02. This margin of error tells us that, with a probability of 95%, the point estimate of 0.31 for the proportion agreeing falls within a distance of 0.02 of the actual population proportion agreeing. In other words, the error we will make in predicting the proportion for the entire population using the point estimate of 0.31 from the GSS sample is no greater than 0.02.
b. The 95% confidence interval is

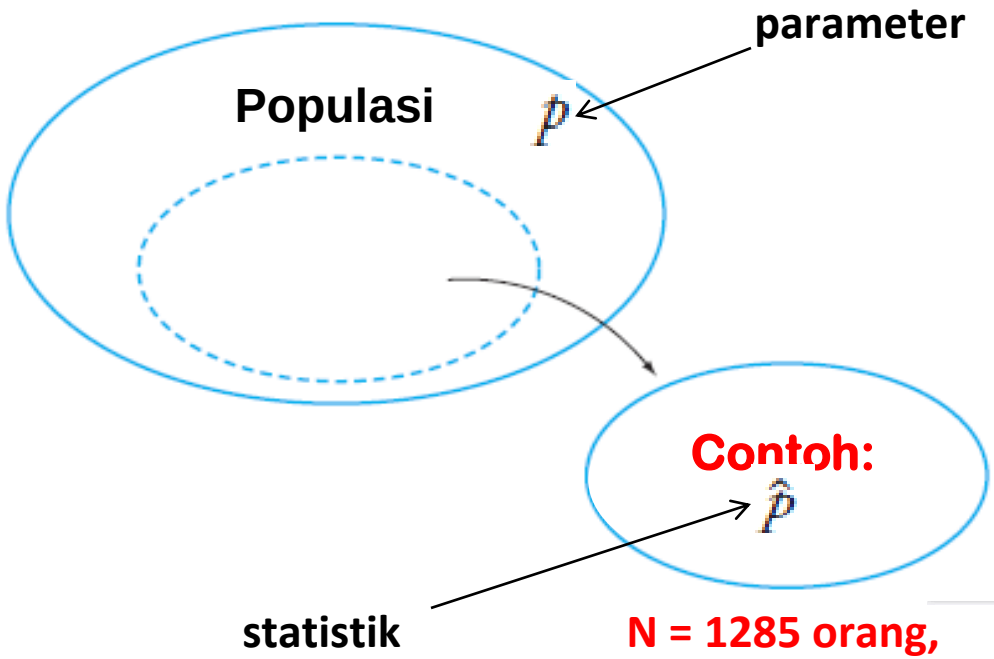
Sample proportion ± 1.96(standard deviation),
which is 0.31 ± 1.96 (0.01), or 0.31 ± 0.02.

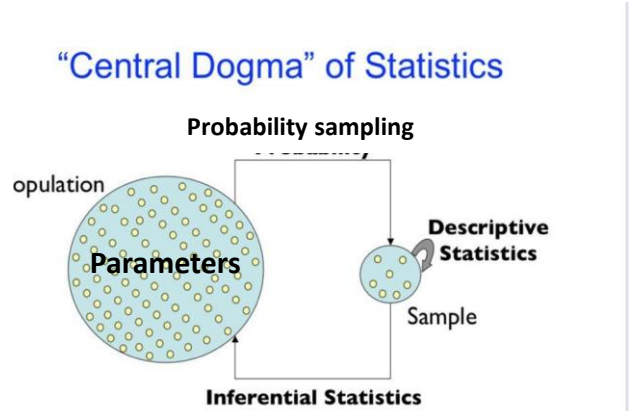
This gives the interval of proportions from 0.29 to 0.33, denoted by (0.29, 0.33). In summary, using this 95% confidence interval, we predict that the population proportion who believed it is much better for everyone involved if the man is the achiever outside the home, and the woman takes care of the home and family was somewhere between 0.29 and 0.33.

Insight

In 1977, when this question was first asked on the GSS, the point estimate was 0.66, and the 95% confidence interval was (0.64, 0.68). The proportion of Americans who agree with this statement has decreased considerably since then.

► *Try Exercise 8.8*





N = 1285 orang,
31% setuju.
Jadi $\hat{p}$ = 0.31

Berapa $\hat{p}$???

MoE ???

SK 95% bagi *p* ???

Selang Kepercayaan

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

Example 2

A Wife's Career

Picture the Scenario

One question on the General Social Survey asks whether you agree or disagree with the following statement: "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." In the 2012 GSS, 31% of 1285 respondents agreed. So the sample proportion agreeing was 0.31. From a formula in the next section, we'll see that this point estimate has an estimated standard deviation of 0.01.

Questions to Explore

a. Find and interpret the margin of error for a 95% confidence interval for the population proportion who agreed with the statement.

b. Construct the 95% confidence interval and interpret it in context.

Think It Through

a. The margin of error for a 95% confidence interval for a population proportion equals 1.96 × (standard deviation), or 1.96(0.01), approximately 0.02. This margin of error tells us that, with a probability of 95%, the point estimate of 0.31 for the proportion agreeing falls within a distance of 0.02 of the actual population proportion agreeing. In other words, the error we will make in predicting the proportion for the entire population using the point estimate of 0.31 from the GSS sample is no greater than 0.02.

b. The 95% confidence interval is

Sample proportion ± 1.96(standard deviation),

which is 0.31 ± 1.96 (0.01), or 0.31 ± 0.02.

This gives the interval of proportions from 0.29 to 0.33, denoted by (0.29, 0.33). In summary, using this 95% confidence interval, we predict that the population proportion who believed it is much better for everyone involved if the man is the achiever outside the home, and the woman takes care of the home and family was somewhere between 0.29 and 0.33.

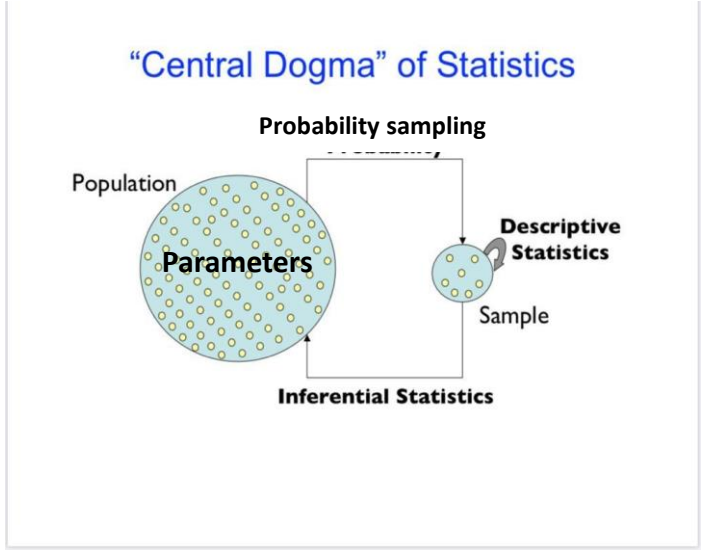
Insight

In 1977, when this question was first asked on the GSS, the point estimate was 0.66, and the 95% confidence interval was (0.64, 0.68). The proportion of Americans who agree with this statement has decreased considerably since then.

► *Try Exercise 8.8*

Pernyataaan: "Sebaiknya biarkan suami sukses berkiprah di luar rumah sedangkan isteri menjaga rumah dan keluarga."disetujui oleh 31% dari 1285 responden.

- Hitunglah dan interpretasikan MoE untuk SK 95% bagi proporsi tsb.
- Susun SK 95% dan interpretasikan hasilnya sesuai konteks.



$$\text{MoE} = 1.96^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96^* \sqrt{\frac{0.31(1-0.31)}{1285}}$$

$$= 1.96^*0.0129 = 0.0253 \approx 0.03$$

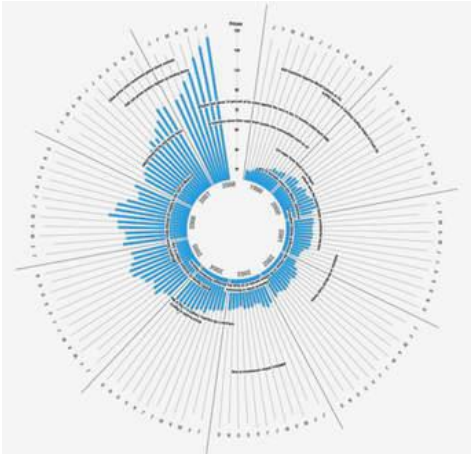
Jadi SK 95% bagi p : 0.31 ± 0.03

atau SK 95% bagi p : (0.28, 0.34)

Kita yakin 95% bhw parameter *p* ada di antara 28% sampai 34%.

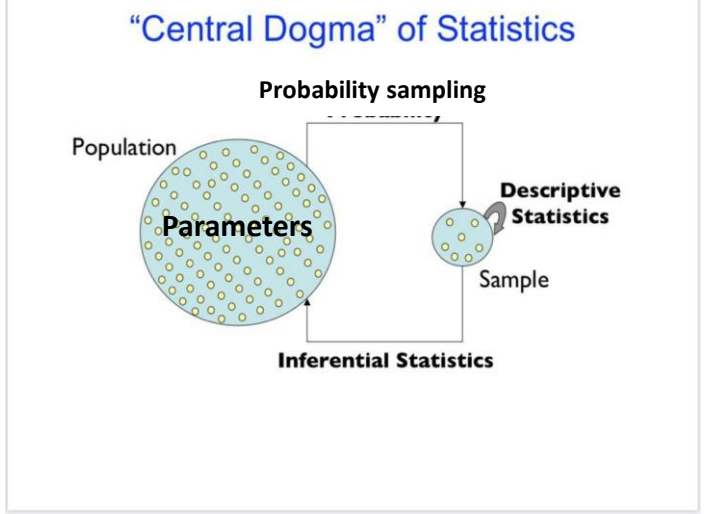
Selang Kepercayaan

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



- SK 95% disusun memanfaatkan titik batas peluang 95% dibawah kurva normal, sehingga diperoleh angka -1.96 di sebelah kiri dan +1.96 di sebelah kanan.
- Luas areal di sebelah kiri dan kanan titik tersebut masing-masing sebesar 2.5% → sering disebut $\alpha = 0.05$
- Jadi SK 95% bisa juga disebut SK (1- α) x 100%
- Dengan mengubah-ubah α maka kita dapat menyusun SK yang lain, misal SK 99% dengan menggunakan $\alpha = 1\%$.
- Perubahan nilai α akan berakibat pada berubahnya nilai 1.96 itu sesuai angka pada tabel sebaran yang ada.
- Misal SK 99% akan menjadi $\hat{p} \pm 2.58 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ ← Selang menjadi lebih lebar

Tabel normal



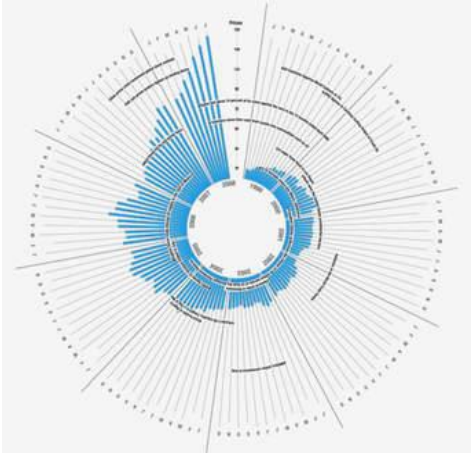
Diskusi Dulu.....

Sesi 2... beres!!!

Sesi 3...

Selang Kepercayaan

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



- INGAT….. SK (1-α)100% sama dengan:

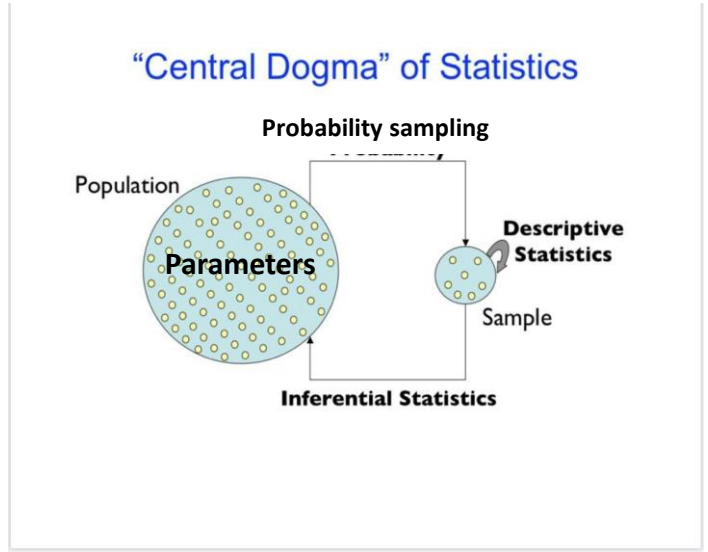
  (penduga titik) ± MoE

- Penduga titik dari parameter μ adalah $\bar{X}$.
- SK (1-α)100% untuk parameter μ adalah

$$\bar{X} \pm z_{\alpha/2}\, \sigma_{\bar{X}}$$

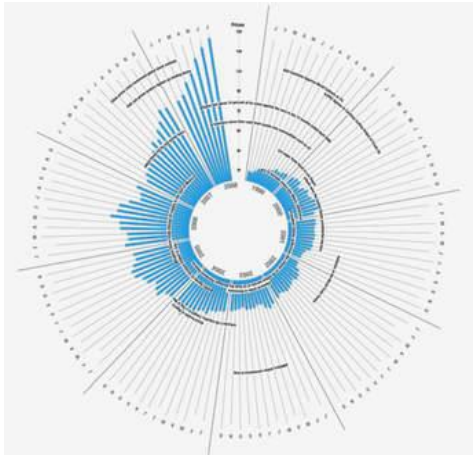
asalkan σ diketahui, dan $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

- Bagaimana jika σ tidak diketahui (lazimnya tdk diketahui)?
- Jika ukuran contoh besar ($n > 30$) maka SK di atas masih benar dan bisa digunakan.
- Jika contoh kecil perlu dimodifikasi karena sebaran dari $\bar{X}$ tidak lagi normal tapi mendekati normal.
- Sebaran ini dikenal sbg sebaran *t-Student*.



Selang Kepercayaan

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

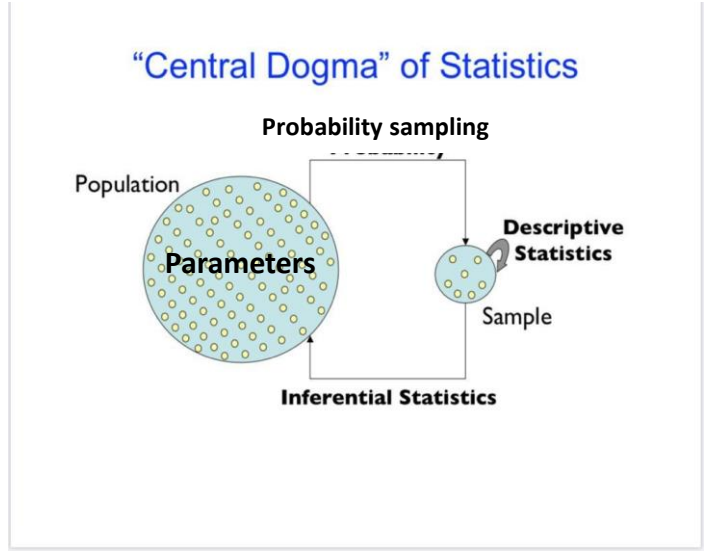


- SK (1-α)100% untuk parameter μ jika σ tidak diketahui dan ukuran *n* kecil adalah

$$\bar{X} \pm t_{\alpha/2(n-1)}\, s_{\bar{X}}$$

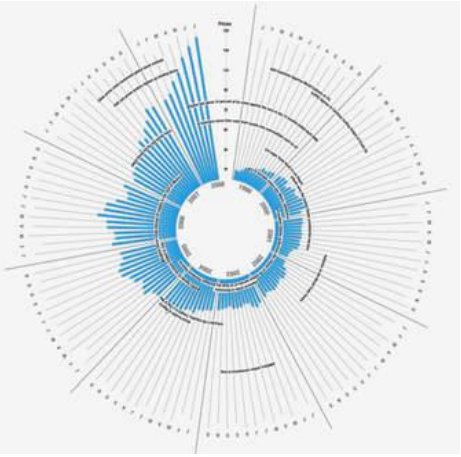
dengan $s_{\bar{X}} = \dfrac{s}{\sqrt{n}}$

- Angka (*n-1*) dikenal sebagai derajat bebas (*degree of freedom*) dari sebaran *t-Student*.



Selang Kepercayaan

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



Construct a 99% confidence interval for the mean daily intake of dairy products for adult men in Example 8.6.

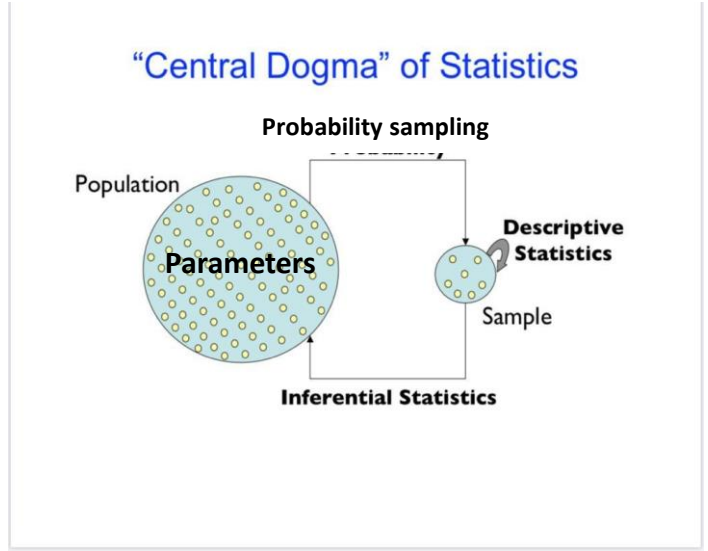
Solution To change the confidence level to .99, you must find the appropriate value of the standard normal z that puts area $(1 - \alpha) = .99$ in the center of the curve. This value, with tail area $\alpha/2 = .005$ to its right, is found from Table 8.2 to be $z = 2.58$ (see Figure 8.10). The 99% confidence interval is then

$$\bar{x} \pm 2.58\left(\frac{s}{\sqrt{n}}\right)$$

$$756 \pm 2.58(4.95)$$

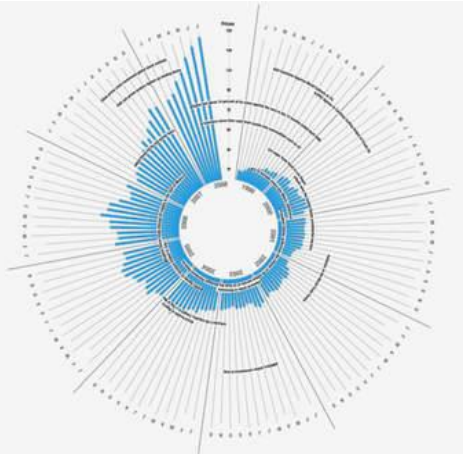
$$756 \pm 12.77$$

or 743.23 to 768.77 grams per day. This confidence interval is *wider* than the 95% confidence interval in Example 8.6.



Selang Kepercayaan

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



Buying on eBay

Picture the Scenario

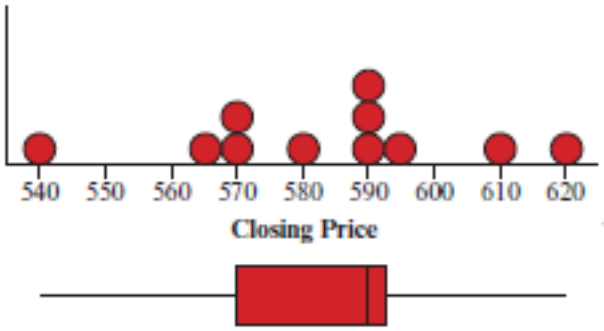
eBay is a popular Internet company for auctioning just about anything. How much can you save by buying items on eBay compared to their actual retail price? Following is a random sample of 11 completed auctions for an unlocked Apple iPhone 5s with 16GB storage in new condition (i.e., item not used, but original packaging might be missing), obtained from eBay in July 2014.

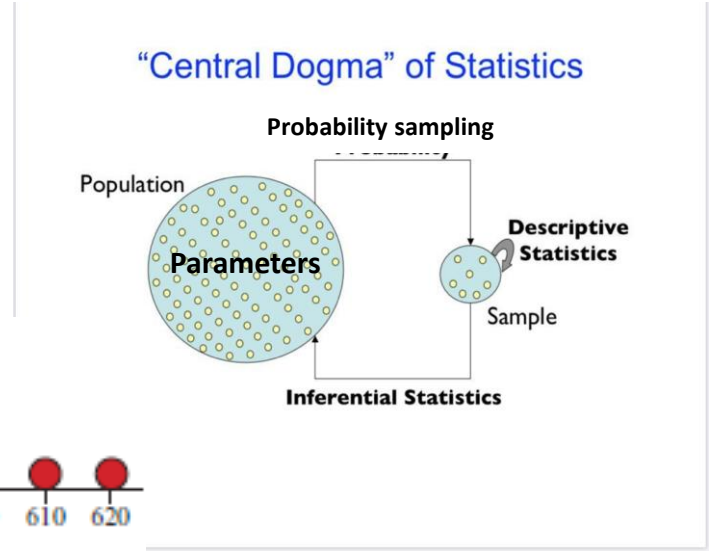
Closing Price (in $): 570, 620, 610, 590, 540, 590, 565, 590, 580, 570, 595

Questions to Explore

a. Use numerical summaries and graphical displays to check the assumptions for using these data to find a 95% confidence interval for the mean closing price on eBay.

b. Find the 95% confidence interval and interpret it. Is there significant savings considering the $649 retail price of the unlocked iPhone 5s in July 2014?





Statistik lima serangkai:

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	11	583.64	22.15	540.00	570.00	590.00	595.00	620.00

Galat baku: $se = s/\sqrt{n} = 22.15/\sqrt{11} = 6.68$

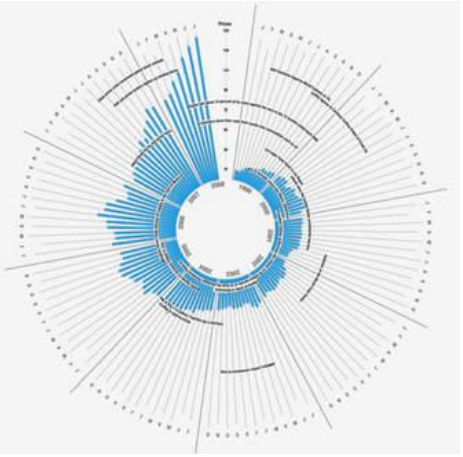
Karena n = 11 maka db utk sebaran t adalah 10. Nilai $t_{0.025(10)}$ = 2.228. Jadi SK 95% bagi μ:

$$\bar{x} \pm t_{.025}(se)$$
$$583.64 \pm 2.228(6.68)$$
$$583.64 \pm 14.88, \text{ or } (568.8, 598.5)$$

Ternyata lebih murah membeli di *eBay*

Selang Kepercayaan

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



Hipotetik

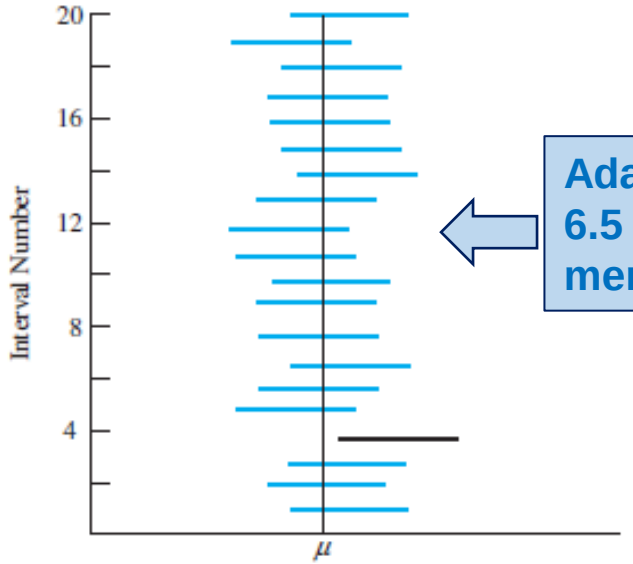
$(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)$

Parameter
$\mu = 6.5$; $\sigma^2 = 8.25$

Contoh ukuran *n=3 → 120* kombinasi

Ada $\frac{10!}{(3!)(7!)} = 120$ kemungkinan

Con-toh	Penga-matan	SK 95% bagi μ
1	2,3,4	(0.516, 5.484)
2	2,3,5	(-0.461, 7.128)
:	:	:
:	:	:
120	9,10,11	(7.516, 12.484)

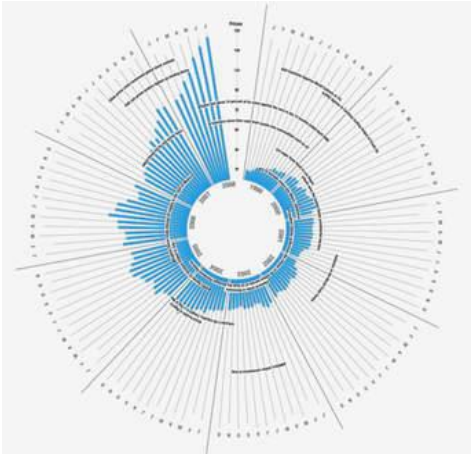


Ada SK yg mencakup μ = 6.5 tetapi ada yang tidak mencakupnya.

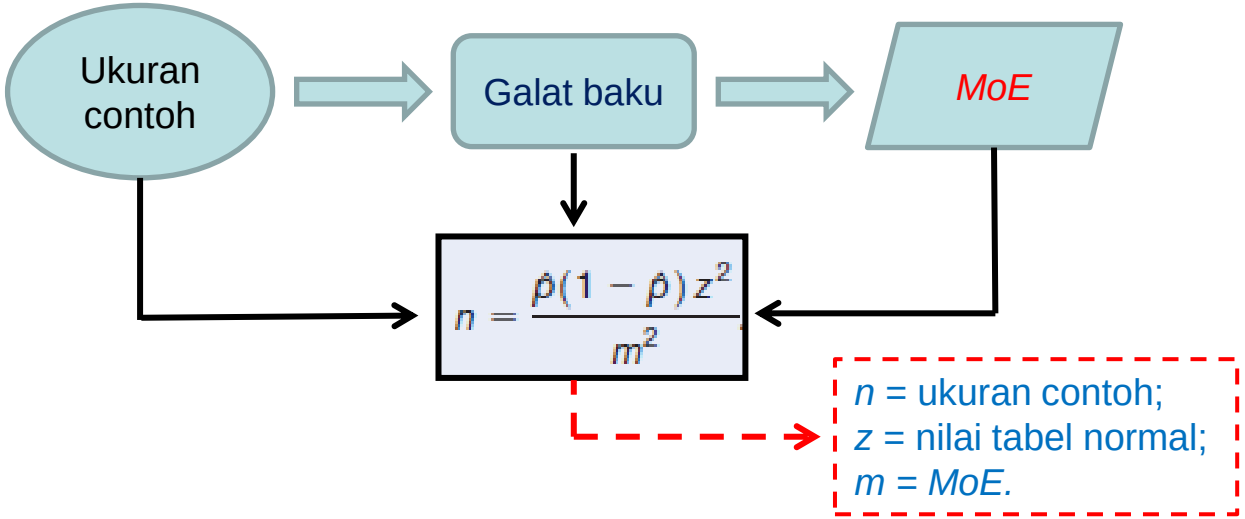
Makna Selang Kepercayaan

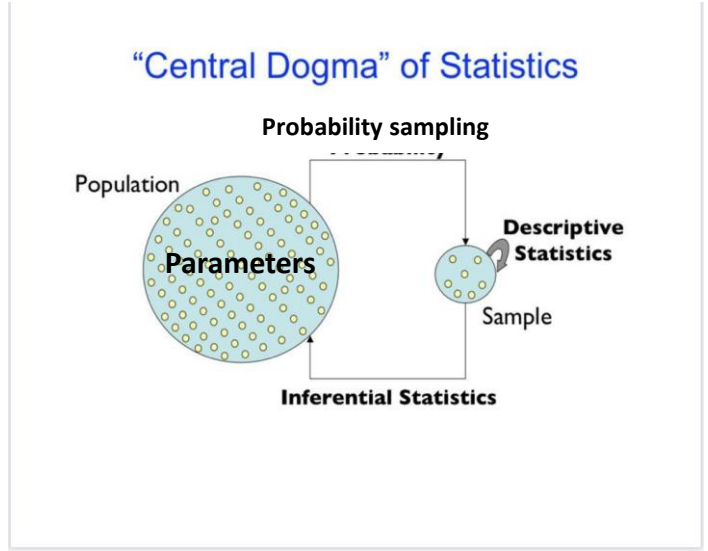
Penentuan Ukuran Contoh

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



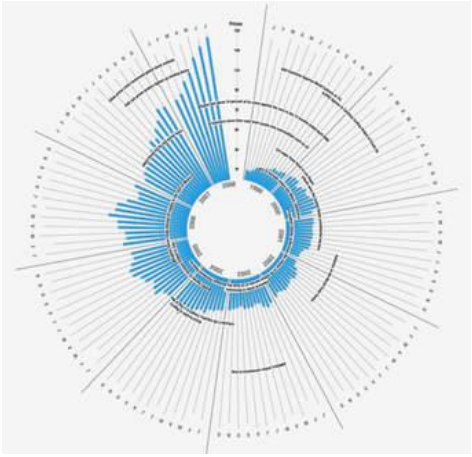
- Ukuran contoh (*n*) ditentukan berdasarkan *MoE* yang ingin ditargetkan.
- Misal ingin *MoE* sebesar 5% atau 1 kg, atau 10 ml, dsb.
- Tentu besarnya *MoE* yg ditargetkan sangat tergantung pada kepentingan studi.
- Prinsipnya spt pada bagan berikut ini (utk parameter *p*):





Penentuan Ukuran Contoh

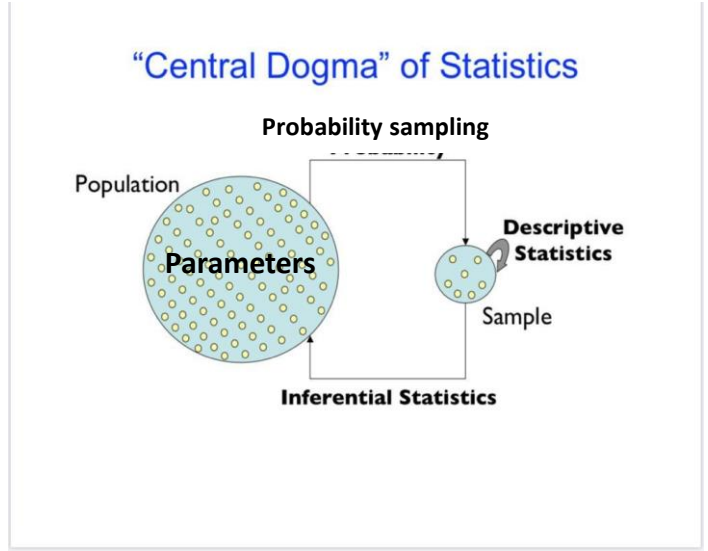
IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



- Ingat bhw → *MoE* = 1.96 * Galat baku
- Atau *MoE* = $1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Kita coba berapa ukuran contoh yang diperlukan jika target *MoE* sebesar 0.04 atau 4%?
- Masukkan ke dalam persamaan di atas sehingga:

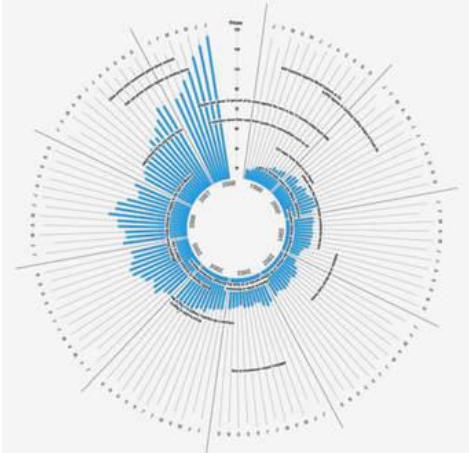
$$0.04 = 1.96\sqrt{\hat{p}(1-\hat{p})/n} \iff n = (1.96)^2\hat{p}(1-\hat{p})/(0.04)^2$$

- Nah, kita butuh $\hat{p}$ padahal data belum dikumpulkan;
- Jika tidak ada informasi tentang $\hat{p}$, gunakan $p = 0.5$ karena nilai ini yang membuat galat baku terbesar.



Penentuan Ukuran Contoh

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



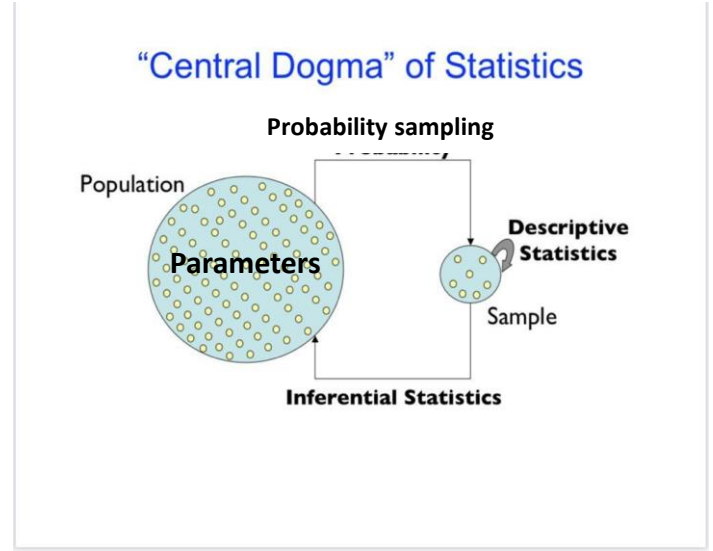
- Jadi jika *MoE* ditargetkan sebesar 4% maka ukuran contoh yang diperlukan adalah:

$$n = (1.96)^2*(0.5)*(1-0.5)/(0.04^2)$$

$$n = 600.25 \approx 600.$$

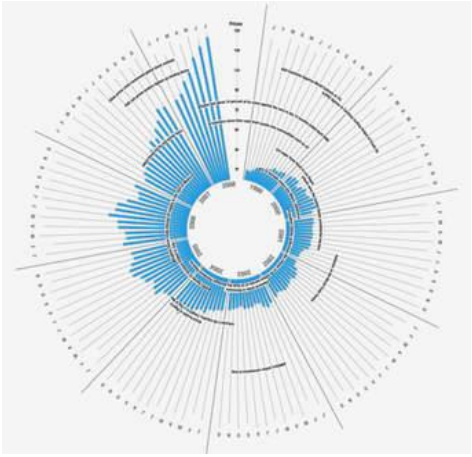
- Ukuran contoh yg diperlukan untuk mencapai target *MoE* sebesar 4% dengan tingkat kepercayaan 95% adalah sebanyak 600 contoh.

- Jika diinginkan tingkat kepercayaan selain 95%, misal 99% maka besaran 1.96 harus disesuaikan.



Penentuan Ukuran Contoh

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



- Bagaimana menentukan ukuran contoh jika parameter yang ingin diduga adalah μ ???
- Prinsipnya sama saja yaitu:

$$MoE = 1.96 * \text{Galat baku } (\bar{X})$$

atau $$MoE = 1.96 * \frac{\sigma}{\sqrt{n}}$$ → σ dari info yg tersedia

- Sehingga ukuran contoh yang diperlukan sebesar:

$$n = \frac{\sigma^2 z^2}{m^2}.$$

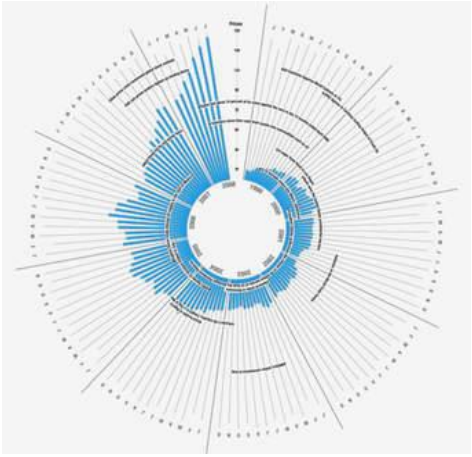
n = ukuran contoh;
z = nilai tabel normal;
m = *MoE*.

How large a sample size is needed so that a 95% confidence interval for the mean number of years of attained education has a margin of error equal to 1 year?

$$n = \frac{\sigma^2 z^2}{m^2} = \frac{3^2(1.96)^2}{1^2} = 35.$$

Peran Komputer

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

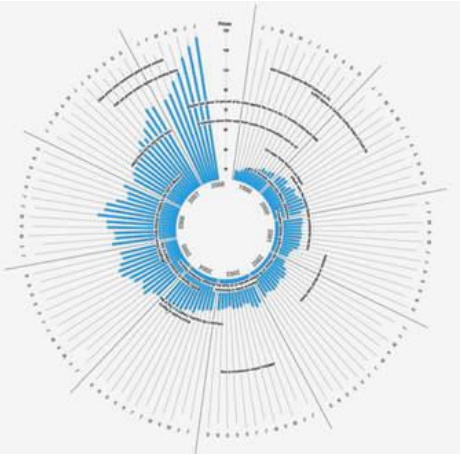


The Bootstrap: Using Simulation to Construct a Confidence Interval

- Bisakah kita menyusun selang kepercayaan dari suatu parameter jika sebaran statistiknya tidak diketahui?
- Misal dari hasil pengukuran bobot badan (dalam *pound*) ingin disusun SK 95% untuk **median** bobot badan, atau **maksimum** bobot badan, atau **simpangan baku** populasi. Maka kita kesulitan mendapatkan sebaran dari statistik (*sampling distribution*) dari **median**, **maksimum** dan **simpangan baku** tersebut.
- Efron (1979) memperkenalkan **metode *bootstrap*** untuk menduga galat baku suatu statistik jika sebaran statistik tidak diperoleh secara matematika. Akibatnya, kita bisa menyusun selang kepercayaan bagi parameternya.

Peran Komputer

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



The Bootstrap: Using Simulation to Construct a Confidence Interval

Perhatikan persoalan menyusun selang kepercayaan dari simpangan baku populasi seperti berikut ini:

- Ingin diketahui seberapa jauh variasi hasil pengukuran suatu timbangan.
- Seseorang menimbang dirinya 10 kali, hasilnya sbb (pound):

  160.2, 160.8, 161.4, 162.0, 160.8, 162.0, 162.0, 161.8, 161.6, 161.8

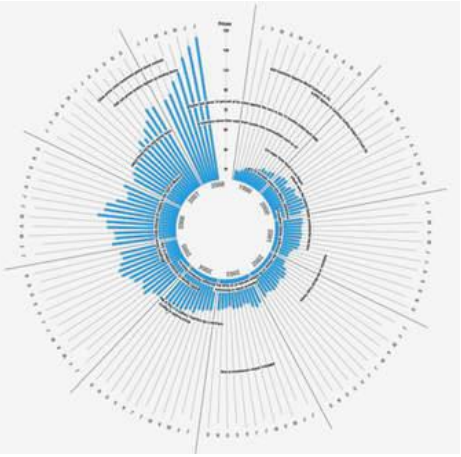
- Ambil secara acak 10 contoh dari 10 pengamatan tersebut dengan pemulihan. Misalkan kita peroleh contoh acak sbb:

  162.0, 160.8, 161.8, 161.8, 162.0, 161.8, 161.8, 162.0, 160.8, 160.8

- Dari 10 contoh acak ini dapat dihitung simpangan bakunya $s = 0.53$

Peran Komputer

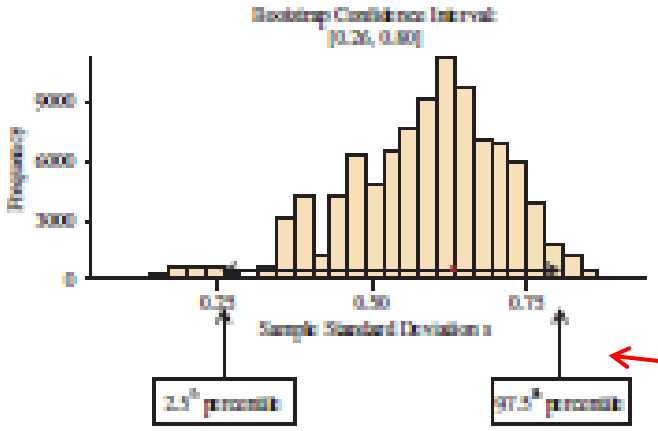
IPB University
Bogor Indonesia
Department of Statistics
Statistics and Data Science Study Program



The Bootstrap: Using Simulation to Construct a Confidence Interval

- **Ulangi** langkah yang sebelumnya sebanyak 100.000 kali, setiap ulangan diperoleh contoh (dengan pemulihan) sebanyak 10 pengamatan dan hitung simpangan bakunya atau s.
- Jadi kita akan memperoleh 100.000 simpangan baku atau s.
- Urutkan nilai s (terkecil ke terbesar) sehingga $s_1 \leq s_2 \leq \cdots \leq s_{100.000}$.
- Selang kepercayaan 95% dari σ adalah nilai $s_{2.500}$ sampai $s_{97.500}$, atau

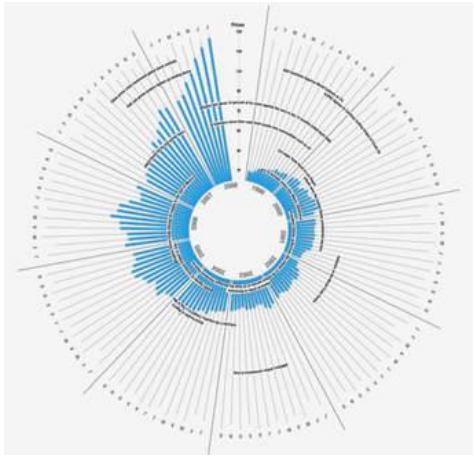
$$s_{2.500} \leq \sigma \leq s_{97.500}$$



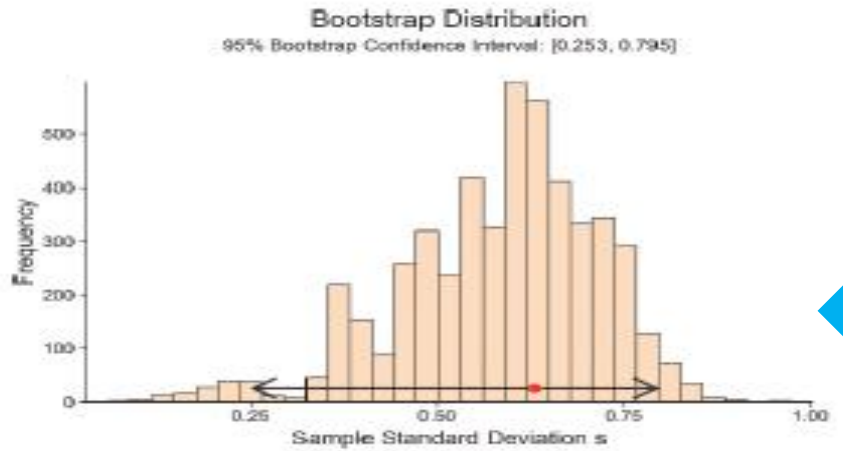
Contoh	Bootstrap samples	*s*
1	160.8, 162.0, …, 160.2	0.53
2	162.0, 161.8, …, 161.6	0.69
:	:	:
100,000	162.0, 160.2, …, 161.4	0.36

Peran Komputer

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



The Bootstrap: Using Simulation to Construct a Confidence Interval



▲ **Figure 8.11** A Bootstrap Frequency Distribution of Standard Deviation Values. These were obtained by taking 100,000 samples of size 10 each from the sample data distribution. **Questions** Does the sampling distribution of the sample standard deviation look approximately normal? What is the practical reason for using the bootstrap method?

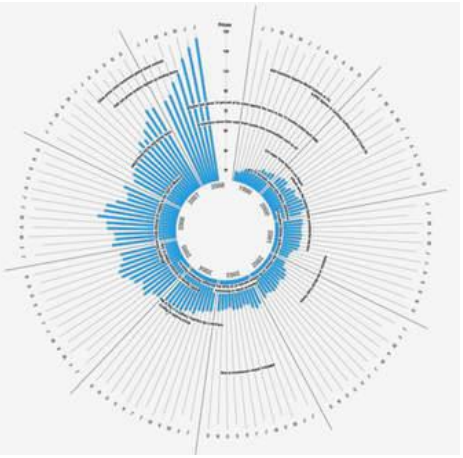
- **Sebaran dari *s* nampak tidak simetrik, agak menjulur ke kiri.**
- **Penggunaan sebaran normal dalam menyusun SK menjadi tidak tepat.**
- **SK yg disusun adalah yang terpendek dari berbagai kemungkinan yg ada.**
- **Teknik bootstrap menjadi cukup sederhana dan menghasilkanSK utk σ,** yaitu (0.253, 0.795)

Setelah dicermati maka selang kepercayaan 95% bagi σ adalah:

$$0.253 \leq \sigma \leq 0.795$$

Peran Komputer

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



The Bootstrap: Using Simulation to Construct a Confidence Interval

INGAT: Teknik *Bootstrap* kita gunakan karena sebaran peluang dari statistik yang menjadi perhatian kita tidak diketahui.
Dalam ilustrasi ini sebaran peluang dari statistik s tidak diketahui.

☺ THANK YOU ☺

IPB University
Bogor Indonesia

Department of Statistics
Study Programs in Statistics and Data Science