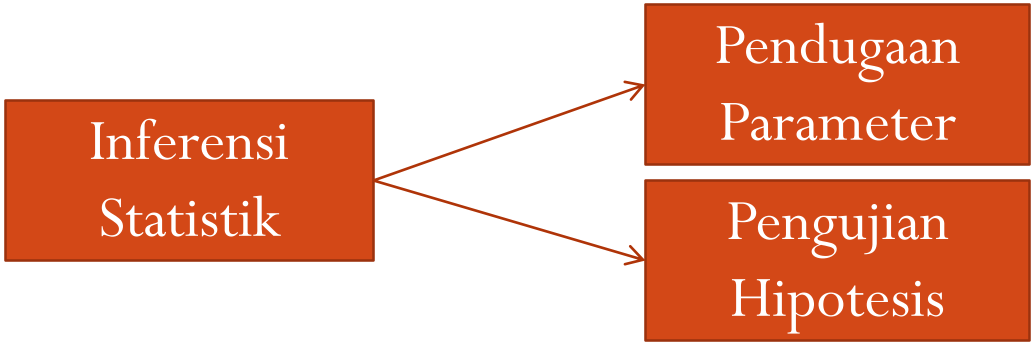
# Pendugaan Parameter

- ***INFERENSI STATISTIK***

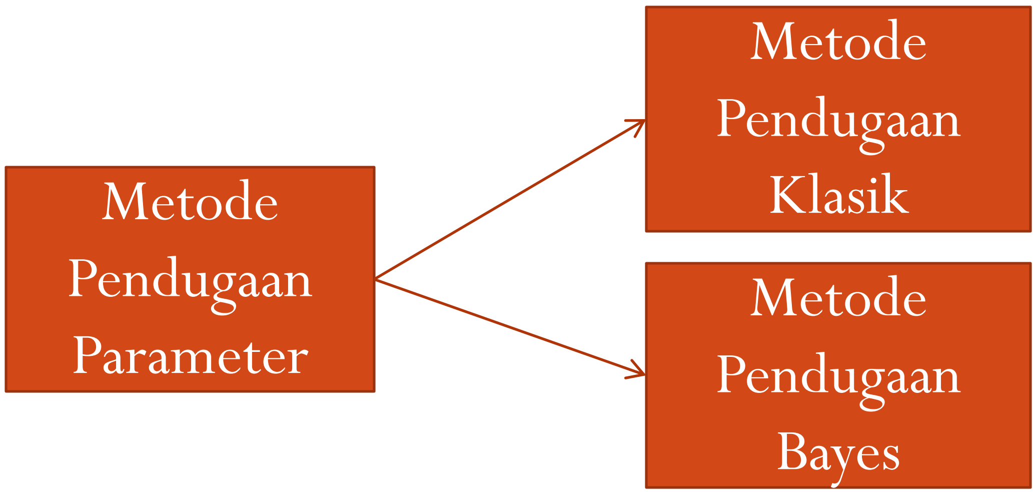
  Inferensi statistik mencakup semua metode yang digunakan dalam penarikan kesimpulan atau generalisasi mengenai populasi.



# Pendugaan Parameter

- Pendugaan parameter berarti melakukan estimasi terhadap nilai dugaan/taksiran suatu parameter tertentu, karena pada umumnya nilai parameter suatu distribusi tidak diketahui
- Contoh :

  Seorang calon dalam suatu pemilihan ingin menduga proporsi yang sebenarnya pemilih yang akan memilihnya, dengan cara mengambil 100 orang secara acak untuk ditanyai pendapatnya. Proporsi pemilih yang menyukai calon tersebut dapat digunakan sebagai dugaan bagi proporsi populasi yang sebenarnya.




- Metode Pendugaan Klasik : Pendugaan dilakukan berdasarkan sepenuhnya pada informasi sampel yang diambil dari populasi.
- Metode Pendugaan Bayes : Pendugaan dengan menggabungkan informasi yang terkandung dalam sampel dengan informasi lain yang telah tersedia sebelumnya yaitu pengetahuan subyektif mengenai distribusi probabilitas parameter.

# Pendahuluan

- $\overline{x}$ digunakan untuk menduga nilai $\mu$ dari suatu populasi
- $S^2$ digunakan untuk menduga nilai $\sigma^2$ dari suatu populasi
- X akan menduga  secara tepat ?
- $\Rightarrow$ Kita menginginkan dist. Sampling memiliki rata-rata yang sama dengan parameter yang diduganya $\rightarrow$ tak bias
- $\Rightarrow$ Untuk sampel yang semakin besar $\rightarrow$ biasnya semakin kecil

# Pendugaan Parameter (Mean)

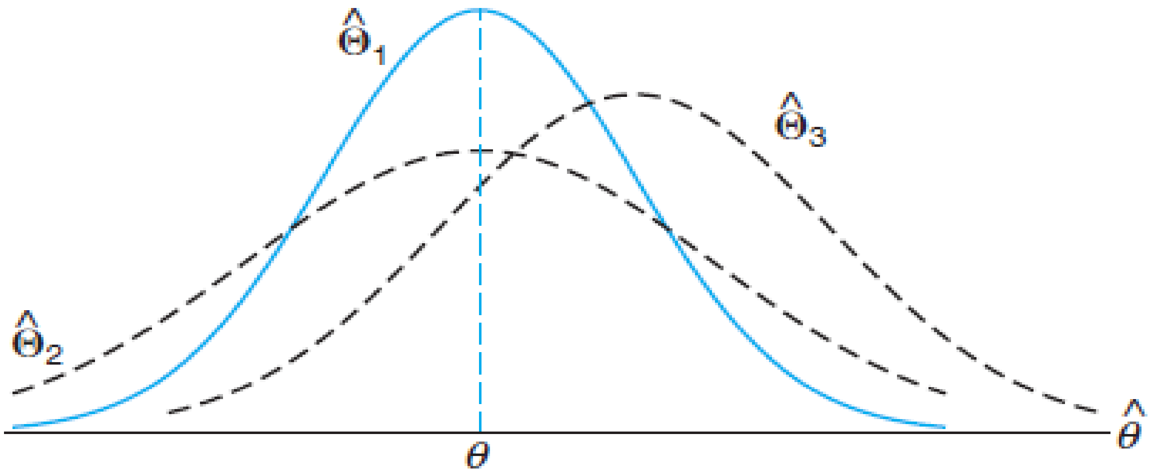



Figure 9.1: Sampling distributions of different estimators of $\theta$.

- $\theta 1$ dan $\theta 2$ adalah penduga tak bias bagi parameter populasi $\theta$ yang sama.
- $\Rightarrow$ pilih penduga yg dist. Samplingnya mempunyai ragam terkecil.
  $\sigma^2_{\theta 1} < \sigma^2_{\theta 2}$ maka $\theta 1$ merupakan penduga $\theta$ yg <u>lebih efisien</u> daripada $\theta 2$.

⇒ Penduga tak bias yang paling efisien sekalipun kecil kemungkinannya menduga parameter populasi secara tepat benar → dugaan selang

$$\theta 1 < \theta < \theta 2 \; ;$$

- $\theta 1$ dan $\theta 2$ tergantung pada nilai statistik $\theta$ suatu sampel dan distribusi sampel bagi $\theta$.
- Karena sampel yang berbeda akan menghasilkan nilai $\theta$ yang berbeda, demikian pula $\theta 1$ dan $\theta 2$, sehingga

  $P(\theta 1 < \theta < \theta 2)$ = nilai tertentu

  $P(\theta 1 < \theta < \theta 2) = 1 - \alpha$
- Untuk $0 < \alpha < 1$ → untuk memperoleh suatu sampel random yang menghasilkan suatu selang mengand. $\theta$ mempunyai peluang $1 - \alpha$

- $(1 - \alpha)$ 100% → selang kepercayaan / interval keyakinan.
- "Semakin panjang SK, semakin yakin bahwa selang tersebut mencakup parameter yang tidak diket. tersebut."

# Pendugaan mean

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

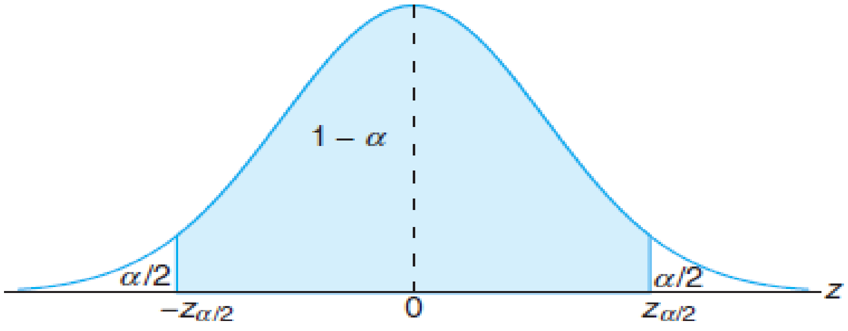
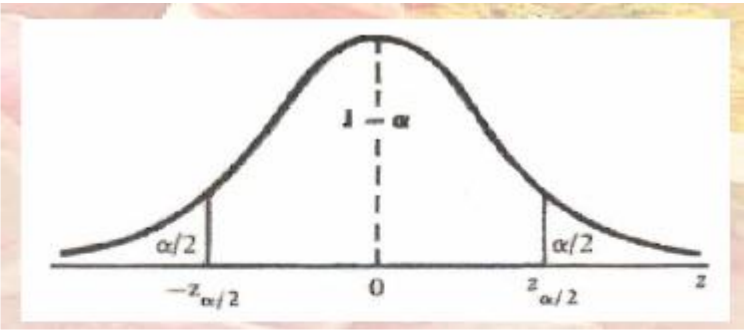



Figure 9.2: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

# PENDUGAAN MEAN




Penduga titik bagi mean populasi μ adalah statistik $\overline{X}$. Bila $\overline{x}$ adalah mean sampel acak berukuran n yang diambil dari suatu populasi dengan ragam $\sigma^2$ diketahui maka selang kepercayaan 100(1-α)% bagi μ adalah

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

CATATAN : Jika $\sigma^2$ tidak diketahui, tetapi sampel berukuran besar ($n \geq 30$), $\sigma^2$ dapat diganti dengan $s^2$.

# Contoh 1 :

- Rataan nilai matematika sampel acak 36 mahasiswa tingkat sarjana adalah 2,6. Hitunglah selang kepercayaan 95% dan 99% untuk rataan nilai matematika semua mahasiswa tingkat sarjana. Anggap bahwa simpangan baku populasinya = 0,3
- Jawab :

$\bar{x} = 2.6$

n = 36 (n $\geq$ 30 → sampel cukup besar)

$\sigma = 0.3$

(a). $(1 - \alpha) = 0.95 \rightarrow \alpha/2 = 0.025$

$z_{\alpha/2} = 1.96$ (Lihat tabel)

$\bar{x} \pm z_{\alpha/2} \quad = 2.6 \pm 1.96 \quad = 2.6 \pm 0.098$

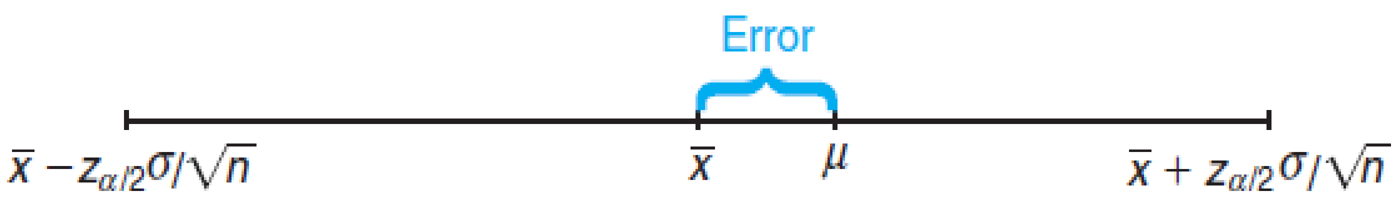
maka SK 95% bagi adalah $2.5 < \mu < 2.70$

- Adapun penduga selang kepercayaan 100(1-$\alpha$)% bagi $\mu$ untuk sampel kecil (n<30); bila $\sigma^2$ tidak diketahui adalah

$$\bar{x} - t_{(n-1,\alpha/2)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1,\alpha/2)} \frac{s}{\sqrt{n}}$$

dengan $t_{(n-1,\alpha/2)}$ adalah nilai t yang luas daerah di sebelah kanan di bawah kurva seluas $\alpha/2$ .

- Bila μ sesungguhnya merupakan nilai pusat selang, maka harapannya $\bar{x}$ menaksir μ tanpa galat. Tetapi, umumnya sampel tidak akan menghasilkan $\bar{x}$ tepat sama dengan μ sehingga taksiran titik umumnya akan meleset (mengandung error). Besarnya galat (error) sama dengan nilai absolut selisih antara μ dengan $\bar{x}$ dengan selang kepercayaan (1-α)100% selisih tersebut akan kurang dari $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.



- Jika ingin diketahui berapa sampel yang harus diambil agar galat dalam menaksir μ akan lebih kecil dari suatu bilangan e yang ditetapkan sebelumnya ➔ maka ukuran sampelnya adalah

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

- Contoh : berapa besar sampel yang diperlukan untuk contoh 1. bila ingin percaya 95% bahwa taksiran untuk μ meleset kurang dari 0,05 ?

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

# PENDUGAAN SELISIH DUA MEAN

- Bila kita mempunyai dua populasi saling bebas dengan mean $\mu_1$ dan $\mu_2$ dan ragam $\sigma_1{}^2$ dan $\sigma_2{}^2$ maka penduga titik bagi selisih antara $\mu_1$ dan $\mu_2$ diberikan oleh statistik $\bar{x}_1 - \bar{x}_2$. Bila $\bar{x}_1$dan $\bar{x}_2$ masing-masing adalah mean sampel acak bebas berukuran $n_1$ dan $n_2$ yang diambil dari populasi dengan ragam $\sigma_1{}^2$ dan $\sigma_2{}^2$ diketahui, maka selang kepercayaan 100(1-$\alpha$)% bagi $\mu_1$-$\mu_2$ adalah

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- dengan $z_{\alpha/2}$ adalah nilai z yang luas daerah di sebelah kanan di bawah kurva normal standard adalah $\alpha/2$ .
- CATATAN : Jika $\sigma_1{}^2$ dan $\sigma_2{}^2$ tidak diketahui, tetapi $n_1$ dan $n_2$ lebih besar dari 30, maka $\sigma_1{}^2$ dan $\sigma_2{}^2$ dapat diganti dengan $s_1{}^2$ dan $s_2{}^2$.

# Contoh :

A study was conducted in which two types of engines, $A$ and $B$, were compared. Gas mileage, in miles per gallon, was measured. Fifty experiments were conducted using engine type $A$ and 75 experiments were done with engine type $B$. The gasoline used and other conditions were held constant. The average gas mileage was 36 miles per gallon for engine $A$ and 42 miles per gallon for engine $B$. Find a 96% confidence interval on $\mu_B - \mu_A$, where $\mu_A$ and $\mu_B$ are population mean gas mileages for engines $A$ and $B$, respectively. Assume that the population standard deviations are 6 and 8 for engines $A$ and $B$, respectively.

***Solution:*** The point estimate of $\mu_B - \mu_A$ is $\bar{x}_B - \bar{x}_A = 42 - 36 = 6$. Using $\alpha = 0.04$, we find $z_{0.02} = 2.05$ from Table A.3. Hence, with substitution in the formula above, the 96% confidence interval is

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}},$$

or simply $3.43 < \mu_B - \mu_A < 8.57$.

# PENDUGAAN SELISIH DUA MEAN

- Adapun penduga selang kepercayaan100(1-$\alpha$)% bagi $\mu_1$-$\mu_2$ untuk sampel kecil; bila $\sigma_1{}^2=\sigma_2{}^2$ tapi nilainya tidak diketahui adalah

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

- dengan derajat bebas untuk distribusi t = v =$n_1$ + $n_2$ – 2 dan

$$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

# contoh

The article "Macroinvertebrate Community Structure as an Indicator of Acid Mine Pollution," published in the *Journal of Environmental Pollution*, reports on an investigation undertaken in Cane Creek, Alabama, to determine the relationship between selected physiochemical parameters and different measures of macroinvertebrate community structure. One facet of the investigation was an evaluation of the effectiveness of a numerical species diversity index to indicate aquatic degradation due to acid mine drainage. Conceptually, a high index of macroinvertebrate species diversity should indicate an unstressed aquatic system, while a low diversity index should indicate a stressed aquatic system.

Two independent sampling stations were chosen for this study, one located downstream from the acid mine discharge point and the other located upstream. For 12 monthly samples collected at the downstream station, the species diversity index had a mean value $\bar{x}_1 = 3.11$ and a standard deviation $s_1 = 0.771$, while 10 monthly samples collected at the upstream station had a mean index value $\bar{x}_2 = 2.04$ and a standard deviation $s_2 = 0.448$. Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.

***Solution:*** Let $\mu_1$ and $\mu_2$ represent the population means, respectively, for the species diversity indices at the downstream and upstream stations. We wish to find a 90% confidence interval for $\mu_1 - \mu_2$. Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07.$$

The pooled estimate, $s_p^2$, of the common variance, $\sigma^2$, is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(11)(0.771^2) + (9)(0.448^2)}{12 + 10 - 2} = 0.417.$$

Taking the square root, we obtain $s_p = 0.646$. Using $\alpha = 0.1$, we find in Table A.4 that $t_{0.05} = 1.725$ for $v = n_1 + n_2 - 2 = 20$ degrees of freedom. Therefore, the 90% confidence interval for $\mu_1 - \mu_2$ is

$$1.07 - (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 1.07 + (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}},$$

which simplifies to $0.593 < \mu_1 - \mu_2 < 1.547$.

# PENDUGAAN SELISIH DUA MEAN

- Selang kepercayaan 100(1-$\alpha$)% bagi $\mu_1$-$\mu_2$ untuk sampel kecil; bila $\sigma_1^2 \neq \sigma_2^2$ tapi nilainya tidak diketahui

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- dengan derajat bebas untuk distribusi t adalah

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}$$

- Bila kita mempunyai dua populasi yang tidak saling bebas (berpasangan), selang kepercayaan 100(1-$\alpha$)% bagi $\mu_D$=$\mu_1$-$\mu_2$ untuk pengamatan berpasangan tersebut adalah

$$\bar{d} - t_{(n-1,\alpha/2)}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{(n-1,\alpha/2)}\frac{s_d}{\sqrt{n}}$$

# contoh

A study was conducted by the Department of Zoology at the Virginia Tech to estimate the difference in the amounts of the chemical orthophosphorus measured at two different stations on the James River. Orthophosphorus was measured in milligrams per liter. Fifteen samples were collected from station 1, and 12 samples were obtained from station 2. The 15 samples from station 1 had an average orthophosphorus content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while the 12 samples from station 2 had an average content of 1.49 milligrams per liter and a standard deviation of 0.80 milligram per liter. Find a 95% confidence interval for the difference in the true average orthophosphorus contents at these two stations, assuming that the observations came from normal populations with different variances.

***Solution:*** For station 1, we have $\bar{x}_1 = 3.84$, $s_1 = 3.07$, and $n_1 = 15$. For station 2, $\bar{x}_2 = 1.49$, $s_2 = 0.80$, and $n_2 = 12$. We wish to find a 95% confidence interval for $\mu_1 - \mu_2$.

Since the population variances are assumed to be unequal, we can only find an approximate 95% confidence interval based on the $t$-distribution with $v$ degrees of freedom, where

$$v = \frac{(3.07^2/15 + 0.80^2/12)^2}{[(3.07^2/15)^2/14] + [(0.80^2/12)^2/11]} = 16.3 \approx 16.$$

Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.84 - 1.49 = 2.35.$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.120$ for $v = 16$ degrees of freedom. Therefore, the 95% confidence interval for $\mu_1 - \mu_2$ is

$$2.35 - 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}} < \mu_1 - \mu_2 < 2.35 + 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}},$$

which simplifies to $0.60 < \mu_1 - \mu_2 < 4.10$. Hence, we are 95% confident that the interval from 0.60 to 4.10 milligrams per liter contains the difference of the true average orthophosphorus contents for these two locations.

# Contoh paired case

A study published in *Chemosphere* reported the levels of the dioxin TCDD of 20 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The TCDD levels in plasma and in fat tissue are listed in Table 9.1.

Find a 95% confidence interval for $\mu_1 - \mu_2$, where $\mu_1$ and $\mu_2$ represent the true mean TCDD levels in plasma and in fat tissue, respectively. Assume the distribution of the differences to be approximately normal.

Table 9.1: Data for Example 9.13

| Veteran | TCDD Levels in Plasma | TCDD Levels in Fat Tissue | $d_i$ | Veteran | TCDD Levels in Plasma | TCDD Levels in Fat Tissue | $d_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 4.9 | −2.4 | 11 | 6.9 | 7.0 | −0.1 |
| 2 | 3.1 | 5.9 | −2.8 | 12 | 3.3 | 2.9 | 0.4 |
| 3 | 2.1 | 4.4 | −2.3 | 13 | 4.6 | 4.6 | 0.0 |
| 4 | 3.5 | 6.9 | −3.4 | 14 | 1.6 | 1.4 | 0.2 |
| 5 | 3.1 | 7.0 | −3.9 | 15 | 7.2 | 7.7 | −0.5 |
| 6 | 1.8 | 4.2 | −2.4 | 16 | 1.8 | 1.1 | 0.7 |
| 7 | 6.0 | 10.0 | −4.0 | 17 | 20.0 | 11.0 | 9.0 |
| 8 | 3.0 | 5.5 | −2.5 | 18 | 2.0 | 2.5 | −0.5 |
| 9 | 36.0 | 41.0 | −5.0 | 19 | 2.5 | 2.3 | 0.2 |
| 10 | 4.7 | 4.4 | 0.3 | 20 | 4.1 | 2.5 | 1.6 |

***Solution:*** We wish to find a 95% confidence interval for $\mu_1 - \mu_2$. Since the observations are paired, $\mu_1 - \mu_2 = \mu_D$. The point estimate of $\mu_D$ is $\bar{d} = -0.87$. The standard deviation, $s_d$, of the sample differences is

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} = \sqrt{\frac{168.4220}{19}} = 2.9773.$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.093$ for $v = n - 1 = 19$ degrees of freedom. Therefore, the 95% confidence interval is

$$-0.8700 - (2.093)\left(\frac{2.9773}{\sqrt{20}}\right) < \mu_D < -0.8700 + (2.093)\left(\frac{2.9773}{\sqrt{20}}\right),$$

# PENDUGAAN PROPORSI

$$\mu_{\widehat{P}} = E(\widehat{P}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

$$\sigma_{\widehat{P}}^2 = \sigma_{X/n}^2 = \frac{\sigma_X^2}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

$$Z = \frac{\widehat{P} - p}{\sqrt{pq/n}},$$

$$P\left(-z_{\alpha/2} < \frac{\widehat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \widehat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha.$$

# *PENDUGAAN PROPORSI*

- Penduga titik bagi proporsi $p$ dalam suatu percobaan binomial diberikan oleh statistik $\hat{p} = X/n$ sedangkan X menyatakan banyaknya keberhasilan dalam n ulangan. Dengan demikian, proporsi sampel $\hat{p} = x/n$ akan digunakan sebagai nilai dugaan titik bagi parameter $p$ tersebut. Bila $\hat{p}$ adalah proporsi keberhasilan dalam suatu sampel acak berukuran n, dan $\hat{q} = 1 - \hat{p}$, maka selang Kepercayaan 100(1-α)% bagi $p$ untuk sampel besar adalah

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- dengan $z_{\alpha/2}$ adalah nilai z yang luas daerah di sebelah kanan di bawah kurva normal standard adalah $\alpha/2$ .

# contoh

Example 9.14: In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.

***Solution*:** The point estimate of $p$ is $\hat{p} = 340/500 = 0.68$. Using Table A.3, we find that $z_{0.025} = 1.96$. Therefore, using method 1, the 95% confidence interval for $p$ is

$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}},$$

which simplifies to $0.6391 < p < 0.7209$.

# PENDUGAAN SELISIH DUA PROPORSI

- Bila $\hat{p}_1$ dan $\hat{p}_2$ masing-masing adalah proporsi keberhasilan dalam sampel acak yang berukuran $n_1$ dan $n_2$ serta $\hat{q}_1 = 1 - \hat{p}_1$ dan $\hat{q}_2 = 1 - \hat{p}_2$ , maka penduga titik bagi selisih antara kedua proporsi populasi $p_1 - p_2$ adalah $\hat{p}_1 - \hat{p}_2$ . Sedangkan selang kepercayaan 100 (1-$\alpha$)% bagi $p_1$ - $p_2$ untuk sampel besar adalah

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- dengan $z_{\alpha/2}$ adalah nilai z yang luas daerah di sebelah kanan di bawah kurva normal standard adalah $\alpha/2$

# contoh

A certain change in a process for manufacturing component parts is being considered. Samples are taken under both the existing and the new process so as to determine if the new process results in an improvement. If 75 of 1500 items from the existing process are found to be defective and 80 of 2000 items from the new process are found to be defective, find a 90% confidence interval for the true difference in the proportion of defectives between the existing and the new process.

***Solution:*** Let $p_1$ and $p_2$ be the true proportions of defectives for the existing and new processes, respectively. Hence, $\hat{p}_1 = 75/1500 = 0.05$ and $\hat{p}_2 = 80/2000 = 0.04$, and the point estimate of $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 = 0.05 - 0.04 = 0.01.$$

$$z_{0.05} = 1.645.$$

$$1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}} = 0.0117.$$

we find the 90% confidence interval to be $-0.0017 < p_1 - p_2 < 0.0217$.

# PENDUGAAN VARIANS

- Dugaan selang $\sigma^2$ ditentukan dengan statistik $X^2 = \frac{(n-1)S^2}{\sigma^2}$.

$$P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha,$$

$$P\left[\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}\right] = 1 - \alpha.$$

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha.$$

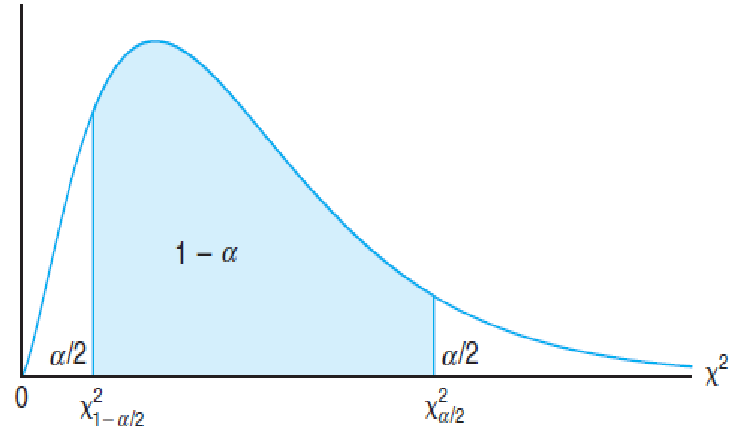



Figure 9.7: $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

where $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ are values of the chi-squared distribution with $n-1$ degrees of freedom, leaving areas of $1-\alpha/2$ and $\alpha/2$, respectively, to the right.

# *PENDUGAAN VARIANS*

- Bila $s^2$ adalah penduga titik bagi varians sampel acak berukuran n yang diambil dari suatu populasi normal dengan varians $\sigma^2$, maka selang kepercayaan 100(1-$\alpha$)% bagi $\sigma^2$ adalah

$$\frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}}$$

- $\chi^2_{(n-1,\alpha/2)}$ adalah nilai $\chi^2$ dengan derajat bebas $v =$ n-1 yang luas daerah di sebelah kanannya sebesar $\alpha/2$

# contoh

- Data berikut menyatakan berat (dalam gr) 10 bungkus bibit sejenis tanaman yang dipasarkan oleh suatu perusahaan : 46,4 ; 46,1 ; 45,8 ; 47 ; 46,1 ; 45,9 ; 45,8 ; 46,9 ; 45,2 dan 46. Carilah selang kepercayaan 95% untuk variansi semua kemasan bibit yang dipasarkan perusahaan tersebut, anggap populasinya berdistribusi normal.
- Jawab :

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

$$= \frac{(10)(21,273.12) - (461.2)^2}{(10)(9)} = 0.286.$$

- Dari tabel v = 9 $\chi^2_{0.025} = 19.023$ and $\chi^2_{0.975} = 2.700$.

$$\frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}}$$

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700}, \rightarrow 0.135 < \sigma^2 < 0.953.$$

# PENDUGAAN RASIO DUA VARIANS

A point estimate of the ratio of two population variances $\sigma_1^2/\sigma_2^2$ is given by the ratio $s_1^2/s_2^2$ of the sample variances. Hence, the statistic $S_1^2/S_2^2$ is called an estimator of $\sigma_1^2/\sigma_2^2$.

If $\sigma_1^2$ and $\sigma_2^2$ are the variances of normal populations, we can establish an interval estimate of $\sigma_1^2/\sigma_2^2$ by using the statistic

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}.$$

According to Theorem 8.8, the random variable $F$ has an $F$-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom. Therefore, we may write (see Figure 9.8)

$$P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha,$$

$$P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha,$$

$$P\left[f_{1-\alpha/2}(v_1, v_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(v_1, v_2)\right] = 1 - \alpha.$$

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(v_1, v_2)}\right] = 1 - \alpha.$$

$$f_{1-\alpha}(\nu_2, \nu_1) = \frac{1}{f_\alpha(\nu_1, \nu_2)}$$

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)\right] = 1 - \alpha.$$

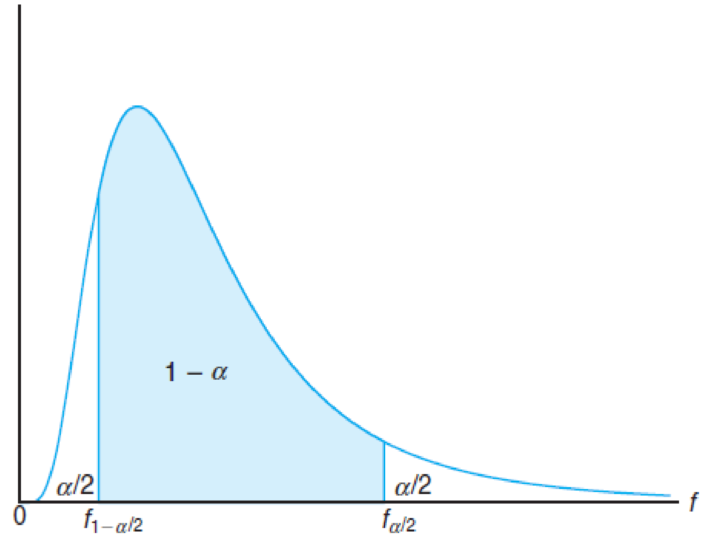



Figure 9.8: $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

# *PENDUGAAN RASIO DUA VARIANS*

- Bila $s_1^2$ dan $s_2^2$ masing-masing adalah varians sampel acak bebas berukuran $n_1$ dan $n_2$ yang diambil dari populasi normal dengan varians $\sigma_1^2$ dan $\sigma_2^2$ , maka penduga titik bagi rasio $\sigma_1^2/\sigma_2^2$ adalah $s_1^2/s_2^2$ , dan selang kepercayaan 100(1-$\alpha$)% bagi $\sigma_1^2/\sigma_2^2$ adalah

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2(v_1,v_2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2(v_2,v_1)}$$

- dengan $f_{\alpha/2(v_1,v_2)}$ adalah nilai *f* untuk derajad bebas $v_1$ dan $v_2$ yang luas daerah di sebelah kanannya sebesar $\alpha/2$ .

# Contoh :

A confidence interval for the difference in the mean orthophosphorus contents, measured in milligrams per liter, at two stations on the James River was constructed in Example 9.12 on page 290 by assuming the normal population variance to be unequal. Justify this assumption by constructing 98% confidence intervals for $\sigma_1^2/\sigma_2^2$ and for $\sigma_1/\sigma_2$, where $\sigma_1^2$ and $\sigma_2^2$ are the variances of the populations of orthophosphorus contents at station 1 and station 2, respectively.

***Solution:*** From Example 9.12, we have $n_1 = 15$, $n_2 = 12$, $s_1 = 3.07$, and $s_2 = 0.80$. For a 98% confidence interval, $\alpha = 0.02$. Interpolating in Table A.6, we find $f_{0.01}(14, 11) \approx 4.30$ and $f_{0.01}(11, 14) \approx 3.87$. Therefore, the 98% confidence interval for $\sigma_1^2/\sigma_2^2$ is

$$\left(\frac{3.07^2}{0.80^2}\right)\left(\frac{1}{4.30}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3.07^2}{0.80^2}\right)(3.87),$$

$$3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991. \qquad 1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Since this interval does not allow for the possibility of $\sigma_1/\sigma_2$ being equal to 1, we were correct in assuming that $\sigma_1 \neq \sigma_2$ or $\sigma_1^2 \neq \sigma_2^2$ in Example 9.12.

# Metoda Penaksiran Bayes

- Bila $\overline{x}$ menyatakan mean sampel acak berukuran n dari suatu populasi normal dengan variansi $\sigma^2$ yang diketahui, dengan distribusi awal dari rataan populasi adalah suatu distribusi normal dengan rataan $\mu_0$ dan variansi $\sigma_0{}^2$, maka distribusi pasca dari rataan populasi juga berdistribusi normal dengan rataan $\mu^*$ dan simpangan baku $\sigma^*$, dengan

$$\mu^* = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{dan} \quad \sigma^* = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$$

$\mu^*$ adalah taksiran bayes dari mean populasi $\mu$

- Selang Bayes $(1-\alpha)100\%$ untuk $\mu$

$$\mu^* - z_{\alpha/2}\sigma^* < \mu < \mu^* + z_{\alpha/2}\sigma^*$$

# Contoh :

- Suatu perusahaan listrik membuat bola lampu yang panjang umurnya berdistribusi hampir normal dengan simpangan baku 100 jam. Dari pengalaman sebelumnya dapat dianggap $\mu$ merupakan suatu nilai peubah acak normal dengan rataan $\mu_0 = 800$ jam dan simpangan baku $\sigma_0 = 10$ jam. Bila sampel acak berukuran 25 bola lampu mempunyai rata-rata umur 780 jam. Cari selang bayes 95% untuk $\mu$.

# Jawab :

- Nilai rataan dari distribusi pasca

$$\mu^* = \frac{(25)(780)(10)^2 + (800)(100)^2}{(25)(10)^2 + (100)^2} = 796$$

- Simpangan baku : $\sigma^* = \sqrt{\frac{(10)^2(100)^2}{(25)(10)^2 + (100)^2}} = \sqrt{80}$

- Selang Bayes 95% untuk μ ➔ $796 - (1{,}96)\sqrt{80} < \mu < 796 + (1{,}96)\sqrt{80}$

$$778{,}5 < \mu < 813{,}5$$

- Selang pendugaan dengan metode klasik

$$780 - (1{,}96)\frac{100}{\sqrt{25}} < \mu < 780 + (1{,}96)\frac{100}{\sqrt{25}}$$

$$740{,}8 < \mu < 819{,}2$$

# Source


- http://www.google.co.id/url?sa=t&rct=j&q=bab-v-pendugaan-parameter.doc&source=web&cd=2&ved=0CCAQFjAB&url=http%3A%2F%2Fbiologiunair.files.wordpress.com%2F2011%2F03%2Fbab-v-pendugaan-parameter.doc&ei=gPbfTpStM5GPiAfBqsyoBQ&usg=AFQjCNF5_7HbTcv4z1lTzCVfCdFXixUZ0A
- Walpole, Ronald E., Myers, Raymond H. 2003. Ilmu Peluang dan Statistik untuk Insinyur dan Ilmuwan, Edisi 6. Bandung: Penerbit ITB.

# SOAL

1. Suatu ujian kimia diberikan kepada 50 siswa wanita dan 75 siswa laki-laki. Siswa perempuan mendapat nilai rata-rata 76 dengan simpangan baku 6, sedangkan siswa laki-laki memperoleh rata-rata 82 dengan simpangan baku 8. Tentukan selang kepercayaan 96% bagi selisih rata-rata nilainya.
2. Dari suatu sampel acak 500 keluarga yang memiliki TV disebuah kota kecil, ditemukan bahwa 340 memiliki TV berwarna. Carilah selang kepercayan 95% bagi proporsi sesungguhnya dari keluarga yang memiliki TV berwarna di kota tersebut.
3. Suatu pengumpulan pendapat umum dilakukan terhadap penduduk kota dan di pinggiran kota untuk menyelidiki kemungkinan didirikannya suatu pabrik kimia. Ternyata 2400 di antara 5000 penduduk kota, dan 1200 di antara 2000 penduduk di pinggiran kota menyetujui rencana tersebut. Buat selang kepercayaan 90% bagi selisih proporsi sebenarnya yang menyetujui rencana tersebut.

# SOAL

4. Seorang peneliti yakin bahwa alat pengukurnya mempunyai simpangan baku $\sigma = 2$. Dalam suatu eksperimen dia mencatat pengukuran 4,1; 5,2; 10,2. Buat selang kepercayaan 90% bagi $\sigma$. Apakah data ini sesuai dengan asumsinya ?
5. Suatu perusahaan taksi ingin menentukan apakah penggunaan ban radial sebagai pengganti ban biasa akan menghemat pemakaian bahan bakar. Dua belas mobil menggunakan ban radial dikendarai melalui jalur pengujian yang telah ditetapkan. Tanpa mengganti pengemudi, mobil yang sama kemudian diberi ban yang biasa dan dikendarai kembali melalui jalan yang sama. Pemakaian bahan, dalam km per liter, tercatat sebagai berikut (Anggap kedua populasi berdistribusi normal) :

| | Kilometer per liter | |
|---|---|---|
| **Mobil** | Ban Radial | Ban Biasa |
| **1** | 4,2 | 4,1 |
| **2** | 4,7 | 4,9 |
| **3** | 6,6 | 6,2 |
| **4** | 7,0 | 6,9 |
| **5** | 6,7 | 6,8 |
| **6** | 4,5 | 4,4 |
| **7** | 5,7 | 5,7 |
| **8** | 6,0 | 5,8 |
| **9** | 7,4 | 6,9 |
| **10** | 4,9 | 4,7 |
| **11** | 6,1 | 6,0 |
| **12** | 5,2 | 4,9 |

- Dengan menggunakan Selang Kepercayaan 95% berapakah selisih jarak tempuh ban radial terhadap ban biasa ?
- Apakah dapat disimpulkan bahwa mobil dengan ban radial menggunakan lebih sedikit bahan bakar daripada ban biasa ?? Jelaskan !!

# SOAL

6. Suatu universitas terkemuka merancang ujian masuk lewat jalur mandiri bagi calon mahasiswanya. Tim penyusun soal tersebut menduga bahwa rata-rata nilai ujian ini akan berbeda dari tahun ke tahun. Berdasarkan pengalaman pada tahun-tahun sebelumnya perbedaan rata-rata nilai ini dinyatakan secara subyektif dengan distribusi data mengikuti normal dengan rataan 72 dan simpangan baku 6.

- Hitunglah peluang awal bahwa rata-rata nilai sesuangguhnya akan berada pada nilai 71,8 dan 73,4 bagi calon mahasiswa tahun depan ?
- Bila soal ujian ini diuji-cobakan pada sampel acak 100 orang calon mahasiswa dan rata-ratanya mencapai 70 dengan variansi 64. Buatlah selang bayes 95% untuk $\mu$ !