Distribusi Sampel

IF2220 Probabilitas dan Statistika
Program Studi Teknik Informatika – STEI – ITB

Bab 8: Distribusi Sampel

01 Statistik Deskripsi Sampel

IF2220 – Probabilitas dan Statistika

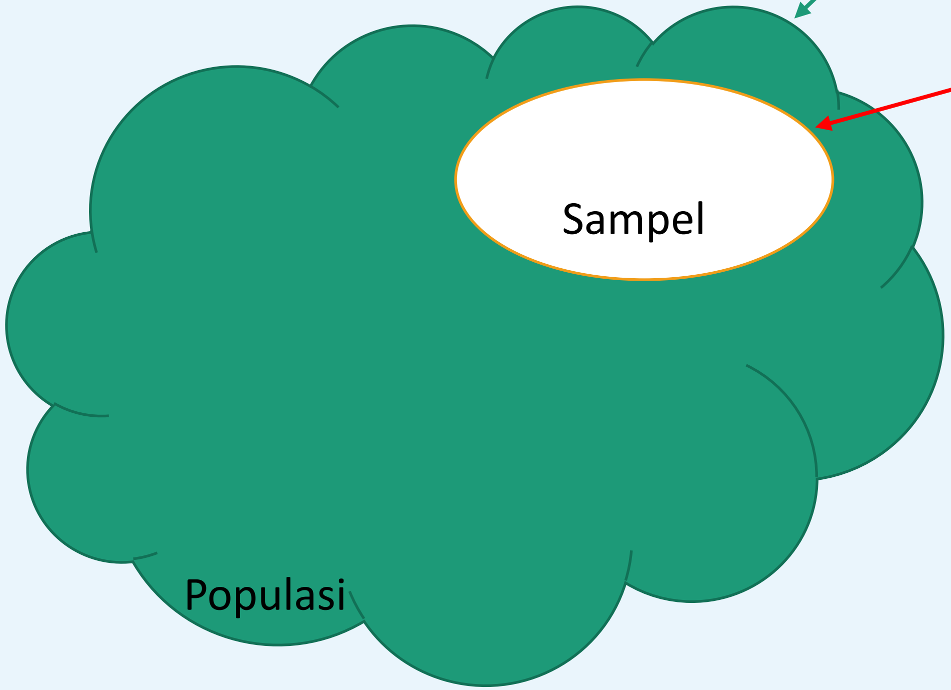
Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

KK IF -Teknik Informatika - STEI ITB

Distribusi Sampel

- Statistik deskripsi sampel
- Distribusi Rataan
- Distribusi Variansi (Chi-square)
- Distribusi T
- Distribusi F

Sampel Acak



Populasi variable random X : keseluruhan pengamatan yang menjadi perhatian.

Sampel : suatu himpunan bagian dari populasi.

n= banyak data sampel/ data set

Biased: prosedur sampling yang menghasilkan inferensi overestimate/underestimate secara konsisten

Sampel Acak [2]

- Definisi statistik dari variabel random : Suatu fungsi (besaran/ukuran) dari suatu sampel acak.
- Secara umum ukuran statistik ada 2:
  1. Ukuran lokasi tengah (center) : rata2, median, modus, dan lainnya.
  2. Ukuran variansi sampel: Jangkauan, variansi sampel, simpangan baku, dan lainnya

Statistik yang Penting

Bila $X_1, X_2, \ldots, X_n$, sampel acak ukuran n:

- **Rataan sampel** dinyatakan oleh statistik:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- **Modus** = nilai X_i yang paling banyak muncul
- **Median**

$$X_{med} = \begin{cases} X_{(n+1)/2} & bila\ n\ ganjil \\ \dfrac{X_{n/2} + X_{(n/2)+1}}{2} & bila\ n\ genap \end{cases}$$

Dimana $X_1, X_2, \ldots, X_n$ diurut membesar

Statistik yang Penting (2)

- **Jangkauan** dari sampel acak $X_1, X_2, \ldots, X_n$ didefinisikan sebagai statistik $J = X_{(n)} - X_{(1)}$, bila $X_{(n)}$ dan $X_{(1)}$ menyatakan masing-masing nilai terbesar dan terkecil dari sampel.
- **Variansi** sampel dinyatakan oleh

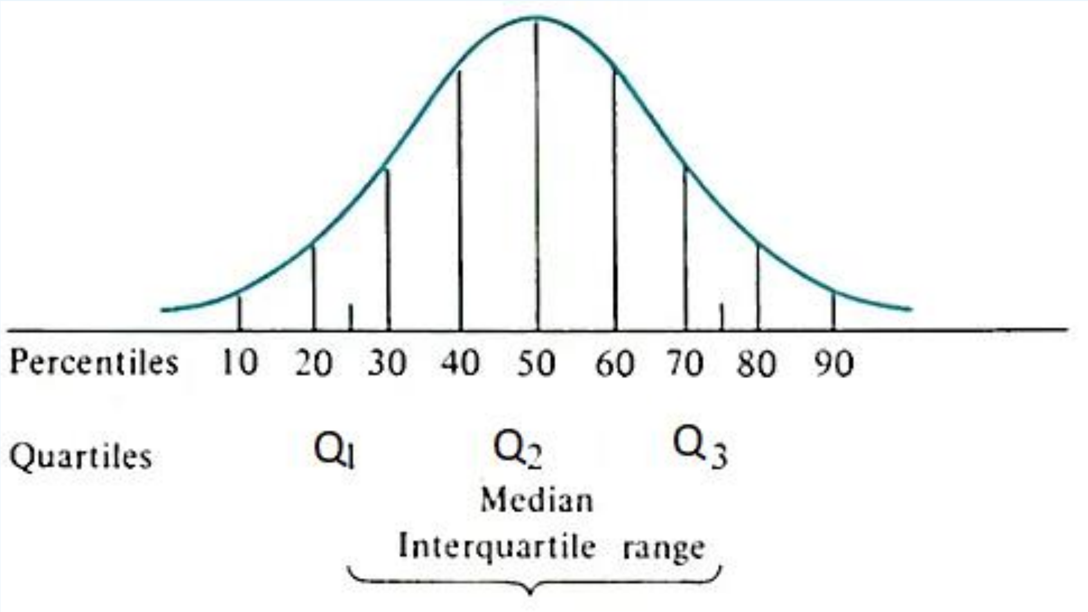
$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

	Sample Size	Mean	Standard Deviation	Variance
Population	N	μ	σ	σ^2
Sample	n	$\bar{x}$	s	s^2

- **Simpangan baku** $S = \sqrt{S^2}$

Statistik yang Penting (2)

- **Persentil rank** , P_{10} = nilai X_i di posisi persentil 10 % dari semua data $X_1, X_2, \ldots, X_n$ diurut membesar.
- **Kuartil**, Q_1, Q_2, Q_3 = nilai X_i di posisi 25 % ,50 %, 75 % dari semua data $X_1, X_2, \ldots, X_n$ diurut membesar. Q_2 = median.
- **Interquartile range**, IQR = $Q_3 - Q_1$



Statistik yang Penting (3)

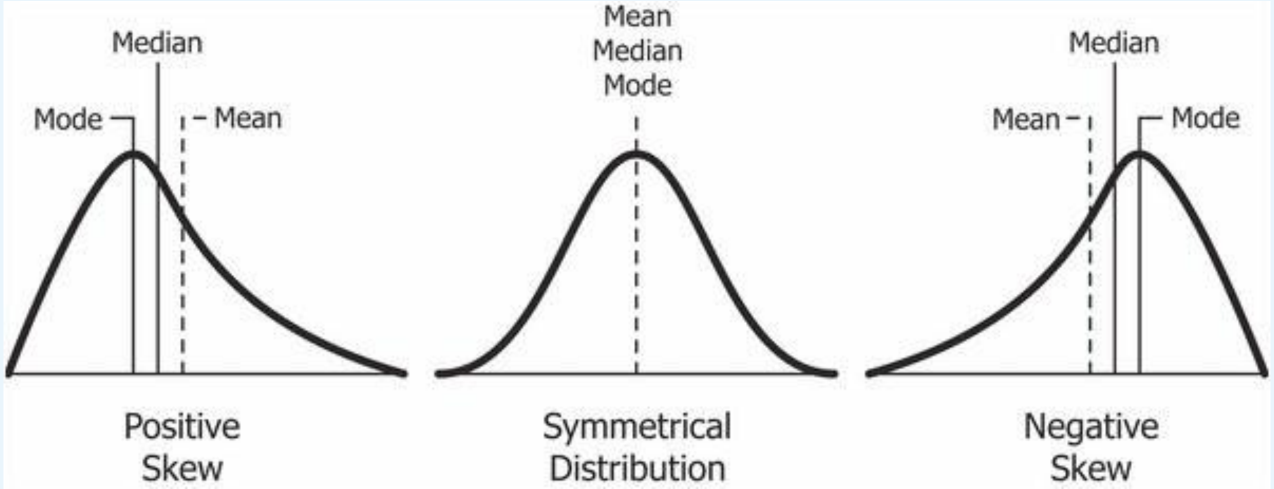
- Nilai Z = posisi nilai X_i thd rataan μ dan simpangan baku σ

$$\boxed{Z = \frac{(X - \mu)}{\sigma}}$$

- Jika nilai X_i > μ maka nilai Z positif dan jika nilai X_i < μ maka nilai Z negatif.

Statistik yang Penting (4)

- **Skewness** (Kemiringan) = ukuran simetri dari fungsi distribusi peluang (fdp), dinyatakan Sk.



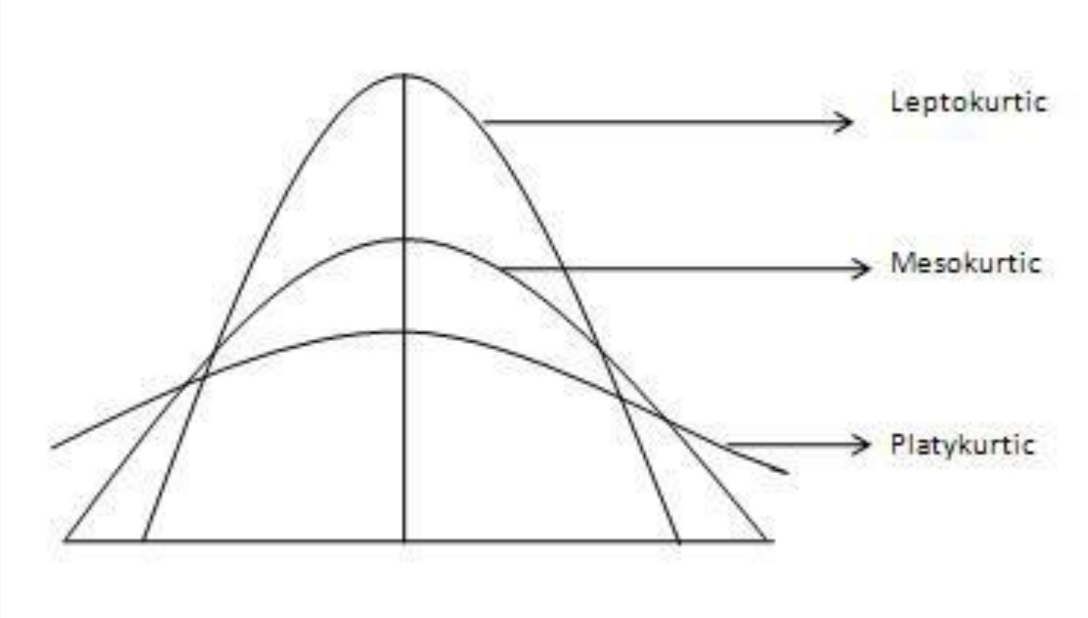
$$Sk = \frac{\{\sum_{i=1}^{n}(X_i - \bar{X})^3\}/n}{(S^2)^{3/2}}$$

Data berdistribusi normal, Skewness = 0

Statistik yang Penting (5)

- **Kurtosis** (keruncingan) = ukuran lancip dari fdp

$$Kur = \frac{(\frac{1}{n}\Sigma(X_i - \bar{X})^4)}{S^4}$$



Data berdistribusi normal,
Kur = 3

02 Distribusi Rataan

IF2222 Probabilitas dan Statistika

Bab 8: Distribusi Sampel

02 Distribusi Rataan

IF2220 – Probabilitas dan Statistika

Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

KK IF -Teknik Informatika - STEI ITB

Distribusi Sampel

- Statistik deskripsi sampel
- Distribusi Rataan
- Distribusi Variansi (Chi-square)
- Distribusi T
- Distribusi F

Distribusi Sampel (1)

- Definisi distribusi sampel : distribusi peluang suatu statistik.
- Definisi distribusi sampel dari rataan , banyak observasi = *n*, diambil dari populasi normal dengan mean = μ dan variansi = σ^2 adalah berdistribusi normal dengan

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

$$\mu_x = \mu_{populasi}$$

$$\sigma_x^{\ 2} = \frac{\sigma_{populasi}^{\ \ 2}}{n}$$

Contoh Distribusi Sampel Rataan

Diketahui data : 3, 2, 3, 2, 3, 4, 4, 2, 3, 4

a) Periksa apakah data berdistribusi normal?

b) Hitung distribusi sampel rataan dari 2 observasi berturutan.

Jawab (1)

a) Rataan = (3+2+3+2+3+4+4+2+3+4)/10 = 3

Median = 3

Modus = 3

Jadi data berdistribusi Normal karena rataan = median = modus

$$S^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}$$

- Variansi = σ^2= (10(96)-900)/10(9) = 0,66
- Data berdistribusi normal dengan μ = 3 dan σ^2 =0,66

Jawab (2)

b) Sampel rataan dari 2 observasi berturutan.

3, 2, 3, 2, 3, 4, 4, 2, 3, 4

No	Nilai
1	(3+2)/2 = 2.5
2	(2+3)/2 = 2.5
3	(3+2)/2 = 2.5
4	(2+3)/2 = 2.5
…	…
8	(2+3)/2=2.5
9	(3+4)/2=3.5

Jawab (3)

- Rataan sampel $\bar{X}$ = (2.5+2.5+2.5+2.5+3.5+4+3+2.5+3.5)/9 = 2,95
- Variansi S^2 = 9(80.75)-(26.5)^2 /(9(8))= 0,34

Bandingkan: μ = 3 dan σ^2 =0,66

Jadi menurut teori, distribusi sampel 2 observasi = berdistribusi normal dengan

- $\bar{X}$ = μ = 3 , dan S^2 = σ^2/n= 0.66/2 = 0,33

Distribusi Sampel

- Teorema Limit Pusat :

Bila $\bar{X}$ rataan sampel acak ukuran n yang diambil dari populasi dengan rataan μ dan variansi σ^2 yang berhingga, maka bentuk limit dari distribusi

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}},$$

bila $n \rightarrow \infty$, ialah distribusi normal baku n(z; 0, 1)

Distribusi Sampel [2]

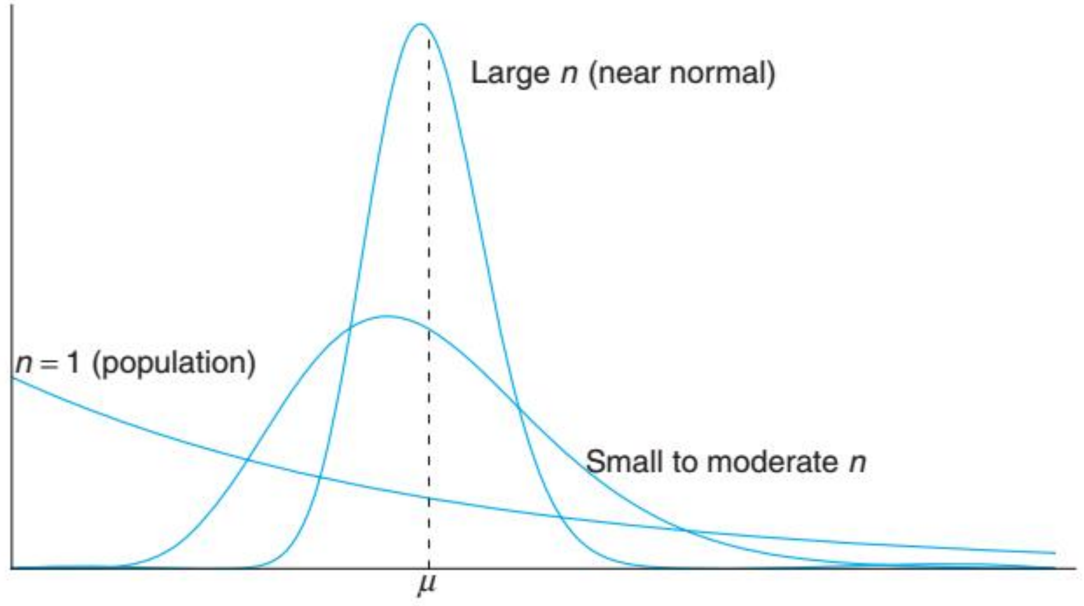


Figure 8.1: Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate n, and large n).

Distribusi Sampel : Example 8. 4

- Contoh soal menghitung peluang sampel random:

Suatu perusahaan memproduksi bola lampu yang umurnya berdistribusi hampir normal dengan rataan 800 jam dan simpangan baku 40 jam.

Hitunglah peluangnya bahwa suatu sampel acak dengan 16 bola lampu akan mempunyai umur rata-rata kurang dari 775 jam.

Distribusi Sampel (2)

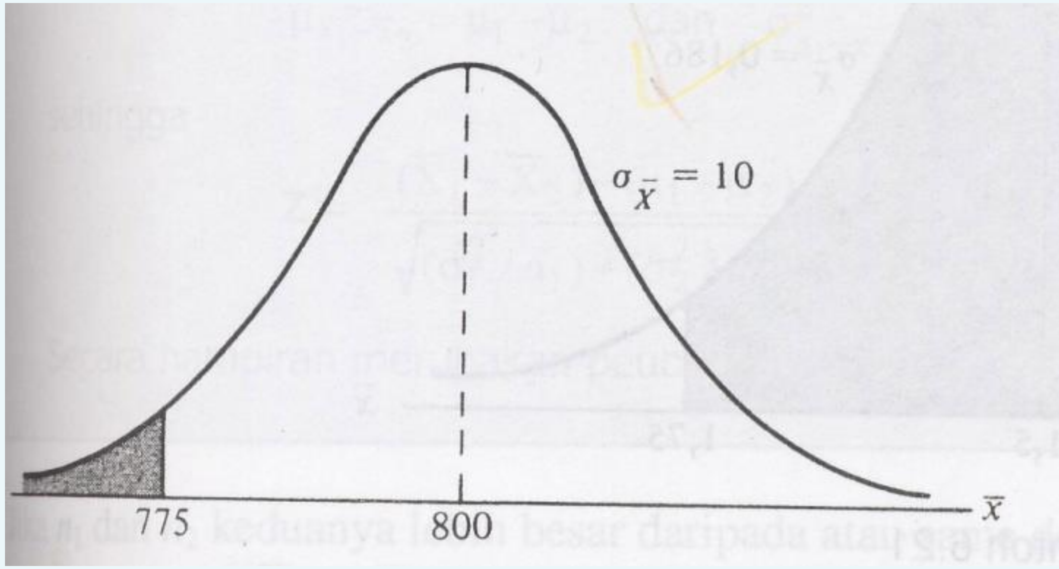
Penyelesaian contoh soal menghitung peluang sampel random :

Secara hampiran, distribusi sampel $\bar{X}$ akan normal dengan $\mu_{\bar{X}}$ = 800 dan $\sigma_{\bar{X}}$ = $40/\sqrt{16}$ = 10. Peluang yang dicari diberikan oleh luas daerah yang dihitami pada gambar. Nilai Z yang berpadanan dengan $\bar{X}$ = 775 adalah

$$z = \frac{775 - 800}{10} = -2{,}5$$

P(X <775) = P(z< -2,5)

= 0,0062



Distribusi Sampel Selisih 2 Rataan

- Bila sampel random ukuran n_1 dan n_2 diambil secara acak dari dua populasi, diskret maupun kontinu, masing-masing dengan rataan μ_1 dan μ_2 dan variansi σ_1 dan σ_2, maka distribusi sampel dari selisih rataan, $\bar{X}_1 - \bar{X}_2$, berdistribusi hampir normal rataan dan variansi diberikan oleh dengan

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2{}_{x_1 - x_2} = \frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}$$

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma^2{}_1}{n_1}\right) + \left(\frac{\sigma^2{}_2}{n_2}\right)}}$$

Secara hampiran merupakan peubah normal baku.

Distribusi Sampel Selisih 2 Mean

Contoh soal menghitung peluang selisih dua rataan dari eksperimen sederhana :

- Suatu sampel berukuran n_1 = 5 diambil secara random dari populasi yang berdistribusi normal dengan rataan μ_1= 50 dan variansi ${\sigma_1}^2$ = 9, dan rataan sampel $\bar{X}_1$ dihitung.
- Sampel random kedua berukuran n_2 = 4 diambil, bebas dari yang pertama, dari populasi lain yang juga berdistribusi normal, dengan rataan μ_2 = 40 dan variansi ${\sigma_2}^2$ = 4, dan rataan sampel $\bar{X}_2$ dihitung.
- Cari $P(\bar{X}_1 - \bar{X}_2 < 8{,}2)$.

Distribusi Sampel Selisih 2 Mean (2)

- Penyelesaian contoh soal menghitung peluang selisih dua rataan dari eksperimen sederhana lebih besar dari nilai tertentu :

Dari distribusi sampel $\bar{X}_1 - \bar{X}_2$ kita ketahui bahwa distribusinya normal dengan rataan

$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$ = 50 – 40 = 10

dan variansi $\sigma^2_{x_1-x_2} = \dfrac{\sigma^2{}_1}{n_1} + \dfrac{\sigma^2{}_2}{n_2}$ = $\dfrac{9}{5} + \dfrac{4}{4} = 2{,}8$

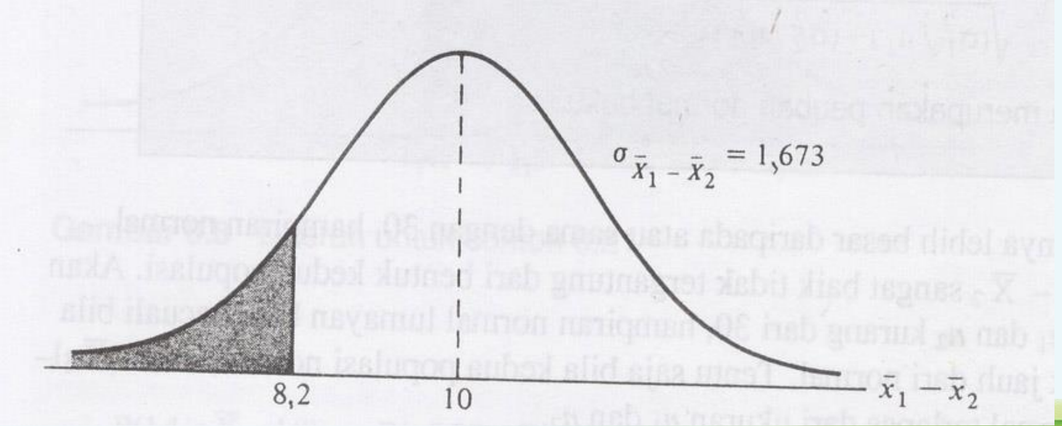
Peluang yang dicari dinyatakan oleh luas daerah yang dihitami dalam gambar.

Berpadanan dengan nilai $\bar{X}_1 - \bar{X}_2$ = 8,2

diperoleh $z = \dfrac{8{,}2 - 10}{\sqrt{2{,}8}} = -1{,}08$

sehingga

$P(\bar{X}_1 - \bar{X}_2 < 8{,}2\ ) = P(Z < -1{,}08)$ = 0,1401



Latihan

- Perusahaan P memproduksi komponen printer dengan rata-rata usia 6,5 tahun dan standar deviasi 0,9 tahun. Perusahaan Q memproduksi komponen tsb dengan rata-rata usia 6,0 tahun dan standar deviasi 0,8 tahun.
- Berapa probabilitas sampel acak berukuran 36 unit komponen dari perusahaan P akan memiliki rata-rata usia minimal 1 tahun lebih panjang daripada rata-rata usia 49 unit komponen dari perusahaan Q?

03 Distribusi Variansi

IF2222 Probabilitas dan Statistika

Bab 8: Distribusi Sampel

03 Distribusi Variansi

IF2220 – Probabilitas dan Statistika

Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

KK IF -Teknik Informatika - STEI ITB

Distribusi Sampel

Statistik deskripsi sampel	Distribusi Rataan	Distribusi Variansi (Chi-square)	Distribusi T	Distribusi F

Distribusi Sampel Variansi

- Bila S^2 variansi sampel acak ukuran n diambil dari populasi normal dengan variansi σ^2, maka statistik

$$\chi^2 = \frac{(\boldsymbol{n}-\mathbf{1})\boldsymbol{S}^2}{\boldsymbol{\sigma}^2}$$

berdistribusi khi-kuadrat dengan derajat kebebasan ***v* = *n* – 1**.

- Distribusi Chi-kuadrat (Tabel A.5) digunakan untuk menghitung peluang variansi sampel.

Distribusi Chi-kuadrat

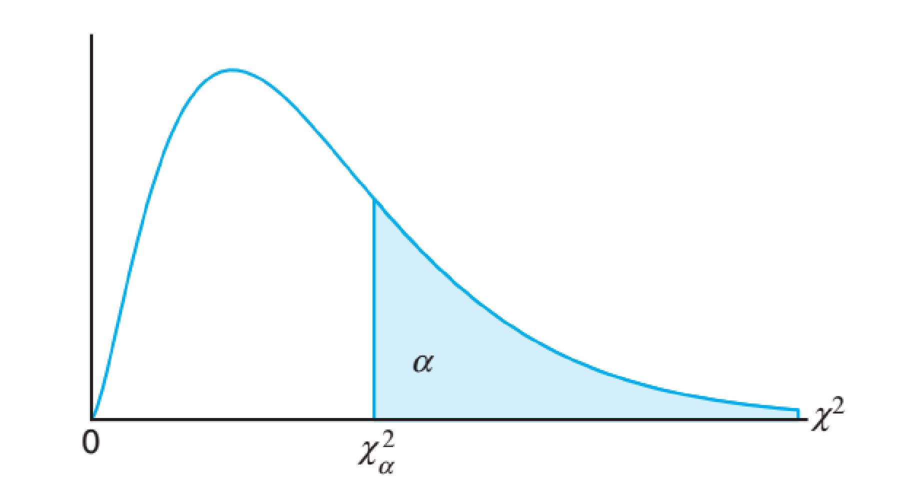


Figure 8.7: The chi-squared distribution.

Table A.5 Critical Values of the Chi-Squared Distribution

	α															
v	0.995	0.99	0.98	0.975	0.95	0.90	0.80	0.75	0.70	0.50	0.30	0.25	0.20	0.10	0.05	0.025
1	$0.0^{4}393$	$0.0^{3}157$	$0.0^{3}628$	$0.0^{3}982$	0.00393	0.0158	0.0642	0.102	0.148	0.455	1.074	1.323	1.642	2.706	3.841	5.024
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386	2.408	2.773	3.219	4.605	5.991	7.378
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366	3.665	4.108	4.642	6.251	7.815	9.348
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357	4.878	5.385	5.989	7.779	9.488	11.143
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351	6.064	6.626	7.289	9.236	11.070	12.832
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348	7.231	7.841	8.558	10.645	12.592	14.449
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346	8.383	9.037	9.803	12.017	14.067	16.013
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344	9.524	10.219	11.030	13.362	15.507	17.535
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343	10.656	11.389	12.242	14.684	16.919	19.023
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342	11.781	12.549	13.442	15.987	18.307	20.483

Distribusi Sampel: Example 8.7

Contoh soal distribusi sampel dari S^2 :

- Suatu pabrik baterai mobil menjamin bahwa baterainya akan tahan rata-rata 3 tahun dengan simpangan baku 1 tahun. Bila 5 baterainya tahan 1,9, 2,4, 3,0, 3,5, dan 4,2 tahun, apakah pembuatnya masih yakin bahwa simpangan baku baterai tersebut 1 tahun ?

Jawab Example 8.7

$$S^2 = \frac{n\sum x_i^{\,2} - (\sum x_i)^2}{n(n-1)}$$

Penyelesaian contoh soal distribusi sampel dari S^2 :

- Mula-mula dihitung variansi sampel :

$$S^2 = \frac{(5)(48{,}26) - (15)^2}{(5)(4)} = 0{,}815 \quad \text{dan} \quad X^2 = \frac{(4)(0{,}815)}{1} = 3{,}26$$

  merupakan suatu nilai distribusi khi-kuadrat dengan derajat kebebasan 4.

- Karena 95 % nilai χ^2 dengan derajat kebebasan 4 terletak antara 0,484 dan 11,143, nilai perhitungan dengan menggunakan $\sigma^2 = 1$ masih wajar, sehingga tidak ada alasan bagi pembuatnya untuk mencurigai bahwa simpangan baku baterainya bukan 1 tahun.

04 Distribusi T dan Distribusi F

IF2222 Probabilitas dan Statistika

Bab 8: Distribusi Sampel

04 Distribusi T dan Distribusi F

IF2220 – Probabilitas dan Statistika

Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

KK IF -Teknik Informatika - STEI ITB

Distribusi Sampel

- Statistik deskripsi sampel
- Distribusi Rataan
- Distribusi Variansi (Chi-square)
- Distribusi T
- Distribusi F

Distribusi-*t* (1)

- Digunakan saat σ tidak diketahui.
- Distribusi suatu statistik dinamakan *T*, dengan

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

- Distribusi *T* untuk menghitung peluang rata-rata untuk banyak sampel sedikit, kecil.

Distribusi-*t* (2)

Misalkan *Z* peubah random normal baku dan *V* peubah random khi-kuadrat dengan derajat kebebasan v. Bila *Z* dan *V* bebas, maka distribusi peubah random *T*, bila $T = \dfrac{Z}{\sqrt{V/v}}$ diberikan oleh

$$h(t) = \frac{\Gamma[\frac{(v+1)}{2}]}{\Gamma(\frac{v}{2})\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}, -\infty < t < \infty,$$

ini dikenal dengan nama distribusi–*t* dengan derajat kebebasan *v*.

Distribusi-t (3)

$t_{1-\alpha} = -t_{\alpha}$

$t_{0.95} = -t_{0.05}$

$t_{0.99} = -t_{0.01}$

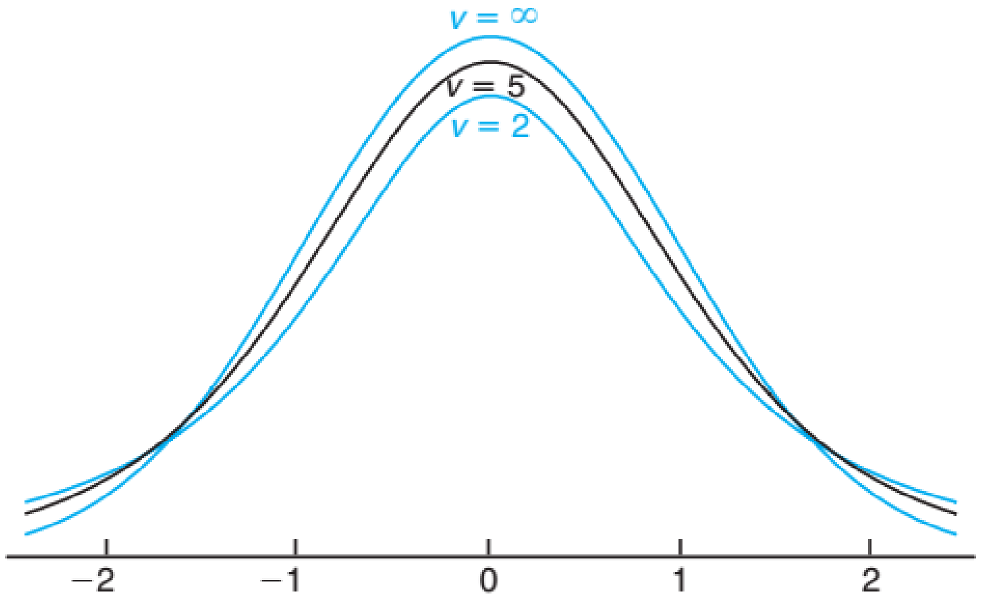


Figure 8.8: The t-distribution curves for $v = 2, 5,$ and ∞.

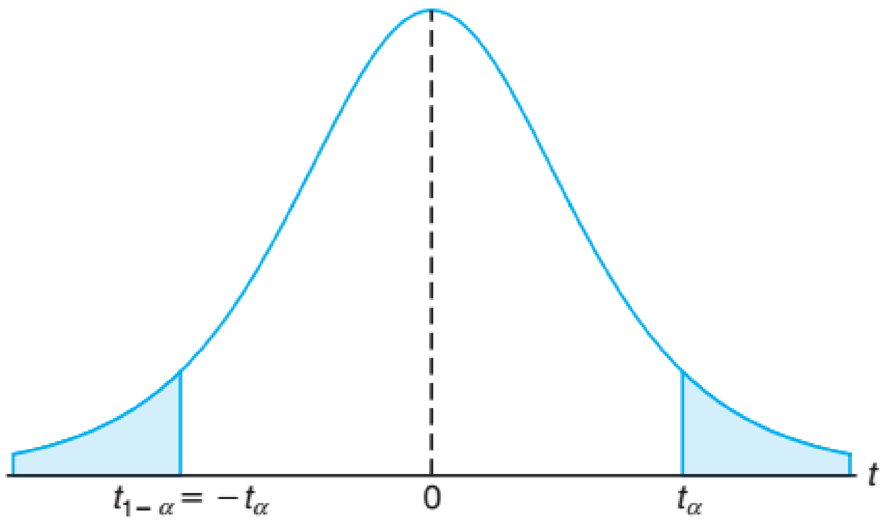


Figure 8.9: Symmetry property (about 0) of the t-distribution.

Distribusi-*t* (4)

Contoh soal distribusi – *t*

- Cari $P(-t_{0,025}<T<t_{0,05})$.

Jawab :

- Karena luas di sebelah kanan $t_{0,05}$ adalah 0,05 dan
- luas di sebelah kiri $-t_{0,025}$ adalah 0,025,
- maka jumlah luas antara $-t_{0,025}$ dan $t_{0,05}$ adalah 1-0,05-0,025 = 0,925

  jadi $P(-t_{0,025}<T<t_{0,05}) = 0,925$

Distribusi-*F* (1)

- Distribusi – *F* didefinisikan sebagai nisbah dua peubah random khi-kuadrat yang bebas, masing-masing dibagi dengan derajat kebebasannya.
- Jadi dapat ditulis

$$F = \frac{U/v_1}{V/v_2}$$

U dan *V* : peubah random bebas.

v_1 dan v_2 : derajat kebebasan.

Distribusi-*F* (2)

- Misalkan U dan V dua peubah random bebas masing-masing berdistribusi khi-kuadrat dengan derajat kebebasan v_1 dan v_2. Maka distribusi peluang acak $F = \dfrac{U/v_1}{V/v_2}$ diberikan oleh

$$h(f) = \begin{cases} \dfrac{\Gamma\left[\frac{(v_1+v_2)}{2}\right]\left(\frac{v1}{v2}\right)^{\frac{v1}{2}} \quad f^{v_1/2-1}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right) \qquad (1+v_1 f/v_2)^{((v_1+v_2)/2)}} & 0 < f < \infty \\ 0 & ,untuk\ nilai\ lainnya \end{cases}$$

Ini dikenal dengan nama distribusi – *F* dengan derajat kebebasan v_1 dan v_2.

Distribusi-*F* (3)

Table A.6 Critical Values of the *F*-Distribution

$f_{0.05}(v_1, v_2)$

v_2 \ v_1	1	2	3	4	5	6
1	161.45	199.50	215.71	224.58	230.16	233.99
2	18.51	19.00	19.16	19.25	19.30	19.33
3	10.13	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	6.26	6.16
5	6.61	5.79	5.41	5.19	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.97	3.87
8	5.32	4.46	4.07	3.84	3.69	3.58
9	5.12	4.26	3.86	3.63	3.48	3.37
10	4.96	4.10	3.71	3.48	3.33	3.22

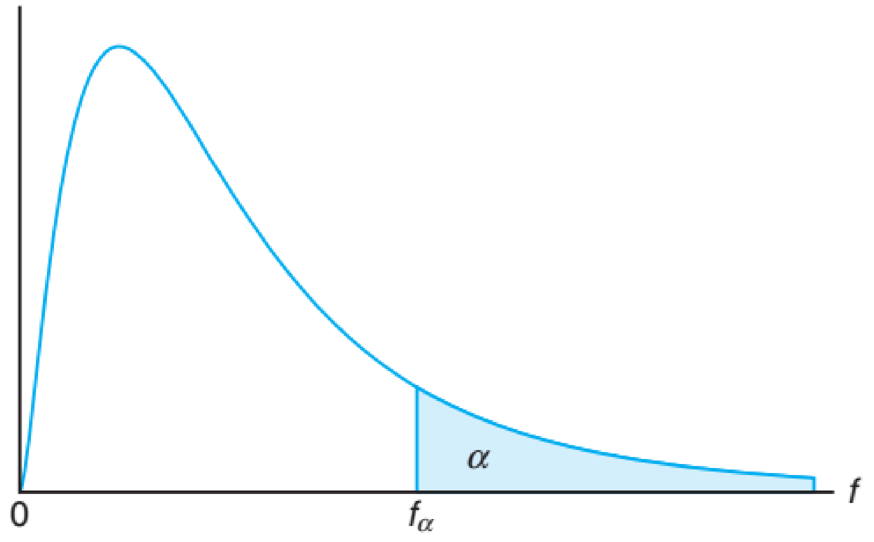


Figure 8.12: Illustration of the f_α for the F-distribution.

Distribusi-*F* (4)

Teorema 8.7:

$$f_{1-\alpha}(v1, v2) = \frac{1}{f_{\alpha}(v2, v1)}$$

- Contoh soal :

  Untuk suatu distribusi – *F* hitunglah $f_{0,95}$ bila $v_1 = 6$ dan $v_2 = 10$

- Jawaban :

  $f_{0,05}\,(10,6) = 4,06.$

  Maka $f_{0,95}\,(6,10) = 1/4,06 = 0,246.$

Distribusi-*F* (5)

- Bila S_1^2 dan S_2^2 variansi sampel random yang bebas ukuran n_1 dan n_2 yang diambil dari dua populasi

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

berdistribusi–*F* dengan derajat kebebasan

$v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$

Distribusi *F* digunakan untuk menghitung peluang rasio 2 variansi sampel.

Latihan

Seorang insinyur mengatakan bahwa suatu proses menggunakan rata-rata 500 gram/ml bahan. Untuk mengujinya, ia menggunakan 25 batch sampel tiap bulan. Jika nilai-*t* yang dihitung antara –t0,05 dan t0.05, ia akan puas.

a) Jika diambil sebuah sampel dengan rata-rata X = 518 gram/ml dan standar deviasi sampel S = 40 gram, hitung nilai t dan probabilitas mendapatkan minimal nilai t tsb.

b) Apakah insinyur tsb layak menyimpulkan prosedur proses tsb baik? Alasannya?

PR

- Bab 8 : #23, 31, 41, 49, 53

Terima Kasih