Distribusi Sampel

Buku Walpole, Bab 8

Populasi dan Sampel

- Populasi (*population*) : kumpulan atau himpunan data yang mendeskripsikan suatu fenomena.

- Sampel (*sample*) : sekumpulan data yang diambil atau diseleksi dari suatu populasi.
- Sampel adalah *himpunan bagian* dari suatu populasi

- Berdasarkan kedua hal tersebut diatas, populasi bisa dipandang sebagai *keseluruhan observasi* yang menjadi perhatian kita.

Populasi

- Jumlah observasi dalam populasi *bisa berhingga* atau *tak berhingga.*
- Terkadang ada berhingga *namun sangat banyak* observasi dalam populasi → diasumsikan *tak berhingga observasi.*

- Distribusi Probabilitas atau Probability Mass Function (PMF) dan Probability Density Function (PDF) merupakan karakteristik dari populasi.
- Expected Value (Mean) dan Variance merupakan parameter dari populasi

Sampel

- Sampel: Himpunan Bagian dari Populasi
- Tujuan Pengambilan Sampel → Statistika Inferensial
  - Kita ingin tahu tentang karakteristik populasi
  - Karakteristik Populasi: PMF, PDF, Mean, Variance, dll.

– Tak mungkin memperoleh karakteristik populasi secara sempurna → tidak mungkin mendapatkan seluruh observasi yg mungkin atau seluruh anggota populasi.

– Solusi: Gunakan sampel untuk memperoleh petunjuk tentang karakteristik populasi.

– Harapan: Sampel bisa secara valid merepresentasikan seluruh populasi.

Mengapa Metode Sampling Penting

Sample → Population

The objective of statistics:
the use of sample information
to infer the nature of a population

Random Sampling

- Pengambilan Sampling tidak boleh Bias => Sample tidak merepresentasikan populasi dengan baik
- Perlu Random Sampling: Jika diinginkan mengambil sampel dengan ukuran sampel *n*:
  - Tiap sampel dengan ukuran sampel *n* (jumlah titik sampel sebesar *n*) memiliki peluang yang sama untuk dipilih dari populasi.

- Katakanlah diambil sampel berukuran 1 dengan melakukan eksperimen 1 kali → 1 peubah acak.
- Distribusi probabilitas dari nilai yang bisa direpresentasikan oleh peubah acak ini → digambarkan oleh PMF (diskret) atau PDF (kontinu) dari populasi

Random Sampling

- Bila eksperimen tidak hanya dilakukan sekali namun n kali → didapatkan n observasi (n realisasi) → sampel berukuran n
- Keluaran pada tiap observasi → direpresentasikan oleh 1 peubah acak: Total ada n peubah acak (n random variables) → $X_1, X_2, \ldots, X_n$

- Distribusi probabilitas (jika diskret) atau pdf (jika kontinu) dari $X_1, X_2, \ldots, X_n$ adalah **identik**:
  - Diambil dari populasi (ruang sampel) yang sama melalui eksperimen yang dilakukan dengan cara yang sama

Random Sampling

- $X_1, X_2, \ldots,$ dan X_n saling independent jika:
  - Random Sampling: Sampel diperoleh secara acak
  - n eksperimen yang dilakukan **tidak ada keterkaitan** antara yg satu dengan yang lain.
  - Keluaran eksperimen ke-i **tidak mempengaruhi** keluaran eksperimen ke-j.

Random Sampling

- Sebuah random sample berukuran n: $X_1, X_2, \ldots, X_n$ dari populasi dengan distribusi probabilitas atau probability density function $f(x)$.

- Bila eksperimen untuk membangkitkan n observasi tsb dilakukan secara random dan dgn prosedur yg sama maka: X_1, X_2, ..., dan X_n akan saling independent dan memiliki distribusi probabilitas atau pdf yg identik (yaitu f(x)) :

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

Tujuan Proses Sampling: Gambaran

- Bayangkan anda ingin mencari proporsi penduduk Indonesia yang lebih menyukai kopi dari pada teh. Dan proporsi ini diberi label P
- Tidak mungkin anda menanyai penduduk Indonesia satu per satu.

- Yang anda lakukan adalah ambil sampel berukuran n, lalu hitung proporsinya yaitu $\hat{p}$.
- $\hat{p}$ belum tentu persis sama dengan P
- $\hat{p}$ umumnya berbeda jika anda mengambil sampel berukuran n yang lain.

- $\hat{p}$ adalah peubah acak (random variable) juga!!
- $\hat{p}$ adalah contoh dari yang disebut dengan **Statistik**.

Statistik

- Setiap fungsi dari seluruh peubah acak (random variables) yang merupakan bagian sebuah random sample disebut **Statistik**.

- Statistik yang mengukur Pusat/Center:
  - Sample Mean
  - Sample Median
  - Sample Mode (Modus)

- Statistik yang mengukur Persebaran:
  - Sample Variance dan Sample Standard Deviation

Sample Mean

- Definisi:
  Jika $X_1, X_2, \ldots, X_n$ merepresentasikan sebuah random sample berukuran n, maka statistik yang merepresentasikan **sample mean** diberikan oleh:

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

- Nilai dari statistik di atas ($\bar{X}$) diberikan oleh:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample Variance

- Jika S^2adalah variance sampel random dengan ukuran n, maka dapat dituliskan:

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]$$

Sampling Distribution

- Statistik adalah fungsi dari beberapa random variable
- Konsekuensi: Statistik sendiri juga merupakan random variable.
- Contoh: Statistik untuk Sample Mean ($\bar{X}$) → fungsi dari X_1, X_2,... , X_n.

- Bila sampel berukuran N yg diambil berbeda2 (mengambil titik sampel yg beda2):
  - Nilai dari $\bar{X}$ juga berbeda2 dan berfluktuasi seiring perubahan nilai X_1, X_2,... , X_n.

Sampling Distribution

- Sample Mean $\bar{X}$ => peubah acak => juga punya distribusi (PDF atau PMF), punya mean, dan punya variance.
- **Distribusi probabilitas (pmf) atau probability density function dari Statistik** (contohnya sample mean) disebut **Sampling Distribution**!

- Sampling Distribution bergantung pada
  - Distribusi dari Populasi
  - Ukuran sampel
  - Cara pengambilan sampel

Sampling Distribution untuk Sample Mean

Anggap sebuah sampel random dengan observasi n diambil dari populasi normal dengan mean μ dan variance σ^2. Setiap observasi $X_i, i = 1,2,\cdots,n$ akan mempunyai distribusi normal sebagaimana dengan populasi yang disampel.

- Bila sample diambil dari populasi yg distribusinya memiliki mean μ_X dan standar deviasi σ_X.

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

Sampling Distribution untuk Sample Mean

- Mempunyai distribusi normal dengan mean:

$$u_{\bar{X}} = \frac{1}{n}\underbrace{(\mu + \mu + \cdots + \mu)}_{n \text{ terms}} = \mu$$

- Variance:

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}\underbrace{(\sigma^2 + \sigma^2 + \cdots + \sigma^2)}_{n \text{ terms}} = \frac{\sigma^2}{n}$$

Sampling Distribution untuk Sample Mean

- Diketahui suatu statistik dijadikan sebagai alat estimasi (estimator) suatu parameter
- Maka statistik tsb dinyatakan sbg **unbiased estimator** bagi parameter yg diestimasi
  - Jika Expected Value dari Statistik sama dengan Parameter yang diestimasi.

- Sample Mean adalah **Unbiased Estimator** bagi True Mean!

Sampling Distribution untuk Sample Mean

Central Limit Theorem

Jika ukuran sampel cukup besar:

- Sample mean dari suatu random sampel akan memiliki sampling distribution yang mendekati distribusi Gaussian (distribusi normal) apapun bentuk distribusi dari populasi yg menjadi sumber pengambilan sampel tadi.

- Makin besar ukuran sampel, pendekatan distribusi normal bagi sampling distribution untuk sample mean makin sempurna.

Central Limit Theorem

- Sebuah peubah acak S adalah jumlahan dari n peubah acak $X_1, X_2, \ldots, X_n$
- Di mana $X_1, X_2, \ldots, X_n$ saling **independent dan memiliki distribusi yg identik** (i.i.d: independent and identically distributed)

- Maka bila n cukup besar (biasanya di atas 30)
- Distribusi dari S sangat mendekati distribusi Gaussian apapun distribusi dari $X_1, X_2, \ldots, X_n$

Ilustrasi Central Limit Theorem

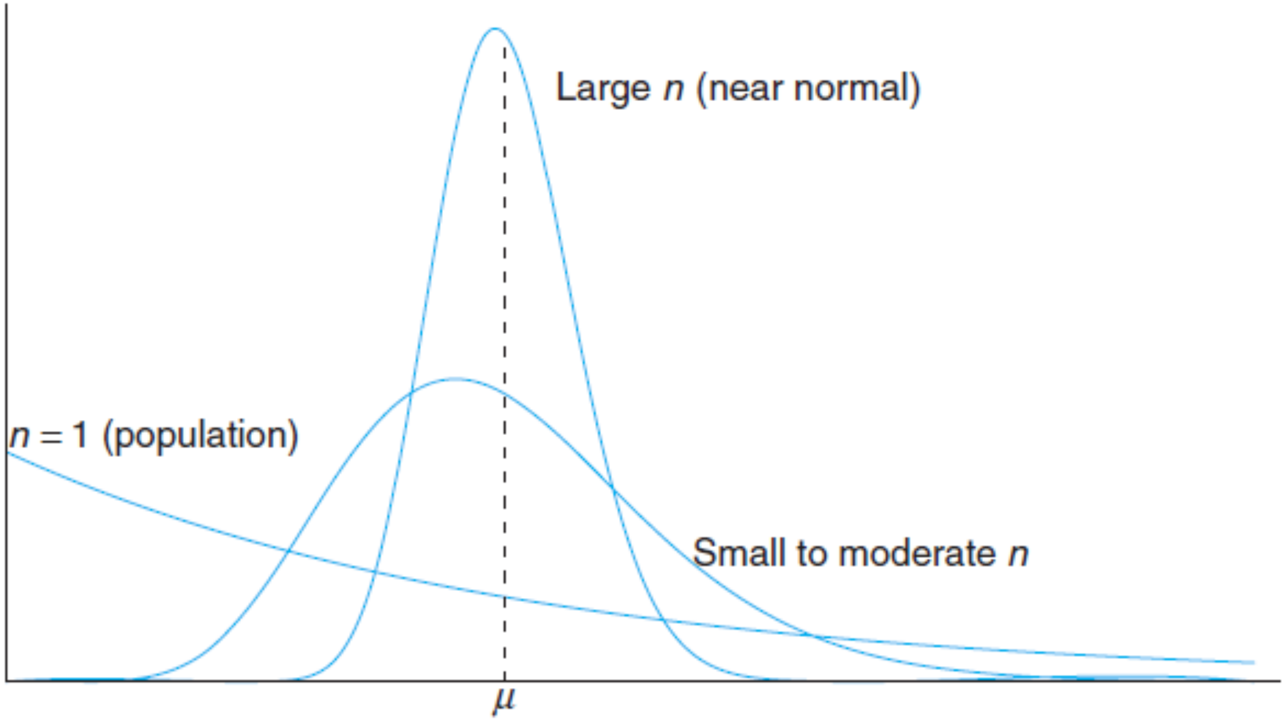


Figure 8.1: Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate n, and large n).

Fig 8.1 of Walpole & Myers (9th edition)

Teorema Central Limit

- Jika $\bar{X}$ adalah rerata sampel random dengan ukuran n yang diambil dari sebuah populasi dengan rerata μ dan variance σ^2, maka batas distribusinya adalah:

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

- Example 8.4 Walpole:

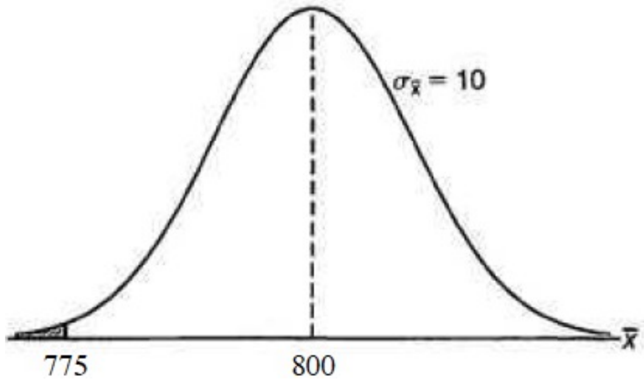
An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Example 8.4 Walpole

Solution:

The sampling distribution of $\overline{X}$ will be approximately normal, with $\mu_{\overline{x}} = 800$ and $\sigma_{\overline{x}} = 40/\sqrt{16} = 10$.

The desired probability is given by the area of the shaded region in Figure.



Corresponding to $\bar{x} = 775$, we find that

$$z = \frac{775 - 800}{10} = -2.5$$

$$P(\overline{X} < 775) = P(Z < -2.5) = 0.0062$$

Sampling Distribution untuk Sample Mean: Penggunaan

- Pada contoh di atas, dicari probabilitas bahwa sample mean lebih kecil dari 775 bila diketahui mean dan standard deviasi dari populasi.

- Penggunaan sering dibalik:
  - Dinyatakan Mean dari Populasi μ=800 (mau **diuji kebenarannya**)
  - Diambil sample berukuran 16 dan dihitung sample mean-nya → ketemu $\bar{X} = 775$.
  - **Rasionalkah** kalau memang Populasi memiliki μ_X=800 dan σ_X=40, maka bisa ditemukan sampel berukuran 16 dgn $\bar{X} = 775$?

Sampling Distribution untuk Sample Mean: Penggunaan

- Kita lalu hitung Probabilitas $P(\bar{X} \leq 775)$ dengan asumsi Mean dari Populasi benar-benar $\mu = 800$
- Diperoleh $P(\bar{X} \leq 775)$ sekitar 0,6% → sangat kecil!
- Kesimpulan: Sangat kecil kemungkinannya jika Populasi memiliki $\mu_X = 800$ dan $\sigma_X = 40$, kita akan temukan sampel berukuran 16 dengan sampel mean di sekitar angka 775.

- Hipotesis bahwa $\mu_X = 800$ jangan-jangan keliru! → mungkin seharusnya kurang dari 800.
- Problem semacam ini disebut **Uji Hipotesis**

Sample Variance

- **Definisi:**
  Jika $X_1, X_2,\ldots, X_n$ merepresentasikan random sample berukuran n, maka statistik yg merepresentasikan **sample** variance adalah

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

- Nilai dari statistik di atas (S^2) diberikan oleh:

$$\boxed{s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)}}$$

Contoh

- Find **the sample variance** of the data 3, 4, 5, 6, 6, and 7, representing the number of trout caught by a random sample of 6 fishermen on June 19, 1996, at Lake Muskoka.

Solution

We find that

$$\sum_{i=1}^{6} {x_i}^2 = 171, \ \sum_{i=1}^{6} x_i = 31, \ n = 6$$

Hence

$$s^2 = \frac{(6)(171) - 31^2}{(6)(5)} = \frac{13}{6}$$

Thus the sample standard deviation

$$s = \sqrt{13/6} = 1.47$$

Sampling Distribution dari Sample Variance S^2

- Teorema: Jika S^2 adalah sample variance dari sebuah random sample berukuran n yg diambil dari populasi berdistribusi Gaussian (Normal) dgn variance σ^2, maka statistik berikut

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

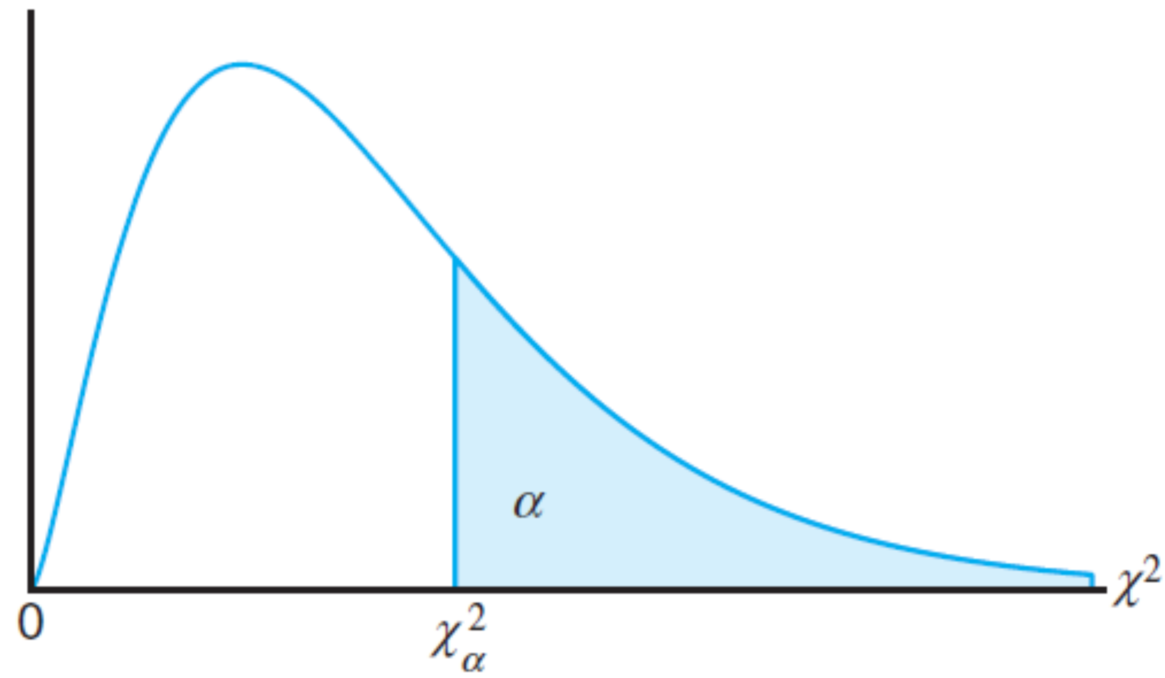
terdistribusi **Chi-Square** dengan ***v*=*n*-1 degree of freedom**.

- Untuk buktikan pernyataan di atas (lihat appendix), perlu diketahui tentang **Peubah Acak Chi-Square**.

Chi-Squared Curve

- The probability that a random sample produces a χ^2-value greater than some specified value is equal to the area under the curve to the right of this value.

- It is customary to let χ_α^2 represent the χ^2 value above which we find an area of α.
- This is illustrated by the shaded region in Figure below



Contoh

- Pabrik pembuat batere mobil menggaransi bahwa batere buatannya akan berumur rata-rata 3 tahun dengan standar deviasi 1 tahun.

- Jika 5 buah batere memiliki usia 1,9; 2,4; 3,0; 3,5; dan 4.2 tahun, apakah pembuat batere bisa yakin bahwa baterenya memiliki standar deviasi 1 tahun?

- Gunakan kriteria berikut sebagai patokan:
  - Apabila

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

berada di **95% area yang dicakup oleh distribusi Chi-Square** maka S^2 yang diperoleh bisa diterima (masuk akal).

- Asumsi: Usia batere mengikuti distribusi normal.

Contoh (Solusi)

- Ingat

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

Pertama-tama kita temukan sample variance:

$$s^2 = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815$$

Kemudian:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(4)(0.815)}{1} = 3.26$$

adalah nilai yang memiliki distribusi chi-squared dengan 4 degrees of freedom.

Contoh (Solusi)

- Dari tabel Chi-Square bisa ditemukan bahwa 95% dari nilai χ^2 dengan 4 degrees of freedom akan jatuh antara 0.484 and 11.143.
- Perhatikan bahwa jika tidak ada informasi lebih detail yang diminta, "**95% area yang dicakup oleh distribusi Chi-Square**" dihitung mulai **tengah area.**

- Dengan demikian, berbasiskan keluaran sampel yang diperoleh, **asumsi bahwa $\sigma^2 = 1$ cukup masuk akal**
- Jadi pabrik tidak perlu khawatir bahwa nilai standard deviasi yang sesungguhnya cukup berbeda dari 1 tahun.

Pengantar Distribusi-t

- Ingat kembali Sampling Distribution untuk Sample Mean
- Ingat Example 8.4 (Contoh tentang Bola Lampu)

- Di situ digunakan Kurva Z untuk berargumen apakah rasional bila Populasi memiliki mean μ_X=800 dan std-deviasi σ_X=40 maka kita dapat menemukan sampel berukuran 16 dgn $\bar{X} = 775$?

- Di situ kita bisa mengasumsikan pernyataan mean μ_X=800 barulah hipotesis yg ingin kita buktikan kerasionalitasnya bila ternyata kita temukan sampel berukuran 16 dgn $\bar{X} = 775$.

Distribusi-t

- Kejanggalan: $\mu_X=800$ barulah hipotesis namun kita menggunakan kurva normal dengan asumsi std-deviasi sudah diketahui pasti yaitu $\sigma_X=40$!!

- Logikanya bila kita belum bisa menghitung μ_X dengan yakin (baru hipotesis) => kecil kemungkinan kita bisa tahu σ_X dengan pasti => karena informasi yg dibutuhkan untuk menghitung 2 besaran tsb sama!

- **Solusi:** Perlu ganti σ_X dengan statistik untuk sample Standard Deviation *S*.

- Problem: Kita belum tentu bisa menggunakan Kurva Standar Normal (karena perlu informasi σ_X) kecuali jika ukuran sampel *n* besar sehingga sample standard deviation s $\approx \sigma_X$. Perlu menggunakan Distribusi-t.

Distribusi-t: Pengujian Sample Mean

- Kita asumsikan random sample dipilih dari suatu populasi normal, akan diperoleh:

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}}$$

- Dengan:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad (1)$$

- Dan

$$V = \frac{(n-1)S^2}{\sigma^2} \qquad (2)$$

Distribusi t: Teorema

- Jika diketahui suatu peubah acak Z adalah peubah acak **normal standar** (μ=0 dan σ=1)
- Dan V adalah peubah acak dengan distribusi **Chi-Squared dengan degree of freedom v**.

- Dan jika Z dan V saling independent, maka bisa didefinisikan peubah acak T sebagai:

$$\boxed{T = \frac{Z}{\sqrt{V/v}}}$$

- Dengan probability density function (pdf) dari t diberikan oleh:

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

- Distribusi dari T di atas dikenal sbg **distribusi-t dengan degree of freedom v**.

Distribusi-t: Pengujian Sample Mean

- Jika $\bar{X}$ adalah sample mean dari sampel yang diperoleh dari populasi dengan distribusi normal dengan mean μ_X dan standar deviasi σ_X, maka Z pada pers(1) jelas terdistribusi standar normal (μ=0 dan σ=1).

- Jika S adalah sample standard deviation dari sampel yang diperoleh dari populasi dengan distribusi normal dengan mean μ_X dan standar deviasi σ_X, maka V pada pers(2) akan terdistribusi Chi-Squared dengan n-1 degree of freedom.
- Untuk jelasnya lihat Slide berjudul "Sampling Distribution dari Sample Variance S^2"

Distribusi-t: Pengujian Sample Mean

- Maka akan tampak jelas bahwa T terdistribusi t dengan n-1 degree of freedom.

- Perhatikan bahwa: **Distribusi t valid untuk T jika** sampel2 (yang kemudian dihitung sample mean-nya sebagai $\bar{X}$ dan sample variance-nya sebagai S^2) diambil random dari **populasi berdistribusi normal**.

- Jika ukuran sample cukup besar ($n \geq 30$), distribusi dari T tidak jauh berbeda dengan standard normal.
- Untuk $n < 30$, sebaiknya dipakai distribusi dari variabel T yang sesungguhnya yaitu distribusi-t dengan n-1 degree of freedom.

Example

- A chemical engineer claims that the population mean yield of a certain batch process is 500 grams per milliliter of raw material.
- To check this claim he samples 25 batches each month. If the computed t-value falls between $-t_{0.05}$ and to $t_{0.05}$, he is satisfied with his claim.

- What conclusion should he draw from a sample that has a mean $\bar{x}$ = 518 grams per milliliter and a sample standard deviation s = 40 grams?
- Assume the distribution of yields to be approximately normal.

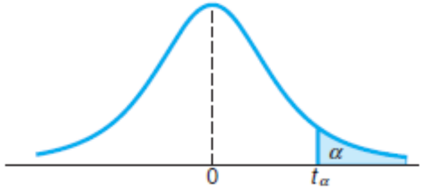


Table A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060

Example: Solution

- From Table A.4 of Walpole, we find that $t_{0.05} = 1.711$ for v= 24 degrees of freedom.
- Hence, the manufacturer is satisfied with his claim if a sample of 25 batches yields a t-value between -1.711 and 1.711.

If $\mu = 500$, then

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2.25 \quad \text{a value well above } 1.711.$$

Hence the manufacturer is likely to conclude that the process produces a better product than he thought.

Distribusi-F (Pengantar)

- Aplikasi distribusi F ditemukan pada permasalahan yang melibatkan 2 atau lebih sampel.
- **Statistik** F didefinisikan sebagai rasio antara dua buah peubah acak yang independent, yang masing-masing memiliki distribusi chi-squared, dan yang masing-masing dinormalisasi/dibagi dengan degree of freedomnya.

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

- Di sini U adalah peubah acak Chi-Squared dengan v_1 degree of freedom
- V adalah peubah acak Chi-Squared dengan v_2 degree of freedom
- U dan V saling independent.

Distribusi-F (Teorema)

- Jika U dan V adalah dua peubah acak independent dengan distribusi chi-squared yang berturut-turut memiliki v_1 dan v_2 degree of freedom.
- Maka probability density function dari:

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

Diberikan oleh:

$$h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0, \\ 0, & f \leq 0. \end{cases}$$

$\Gamma(.)$ adalah Gamma Function yang definisinya bisa dilihat pada slide bagian sebelumnya saat membahas distribusi Chi-Squared

Distribusi dari peubah acak F di atas yang pdf-nya digambarkan oleh h(f) dikenal dengan distribusi-F dengan degree of freedom v_1 dan v_2.

Kurva Distribusi F (Walpole)

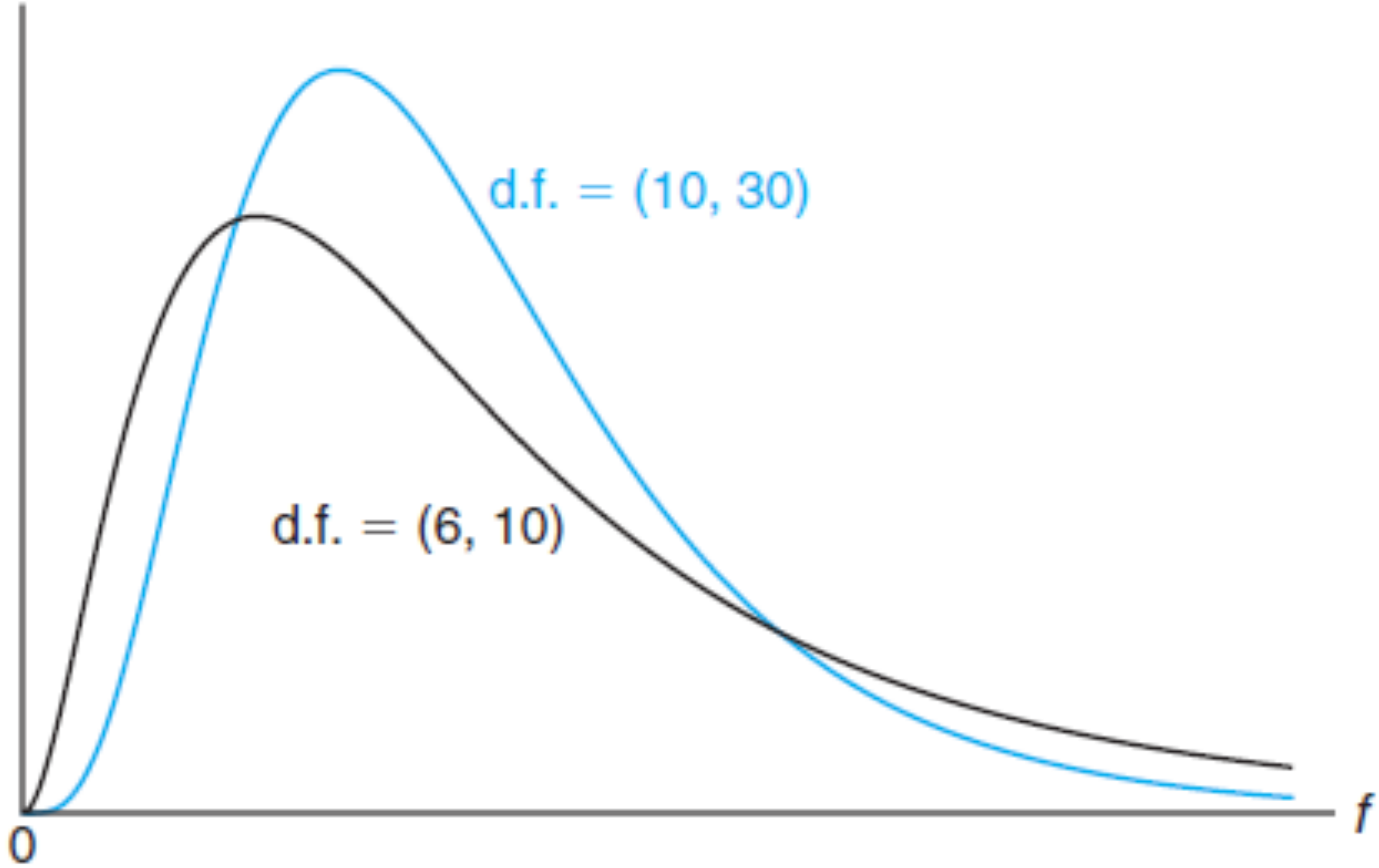


Figure 8.11: Typical F-distributions.

Distribusi-F (Catatan)

- Distribusi F bergantung pada dua degree of freedom yaitu v_1 dan v_2
- v_1 berasal dari peubah acak Chi-Squared pada sisi pembilang (lihat lagi formula dari F: Rasio 2 peubah acak Chi-Squared)
- v_2 berasal dari peubah acak Chi-Squared pada sisi penyebut.

- Dari Kurva Distribusi F, kita akan berhubungan dengan parameter: $f_\alpha(v_1,v_2)$.
- $f_\alpha(v_1,v_2)$: mengindikasikan parameter pada sumbu f pada kurva distribusi-F dengan degree of freedom v_1 dan v_2, sedemikian hingga luasan di bawah kurva distribusi-F di sebelah kanan f_α adalah sebesar α.

Kurva Distribusi F

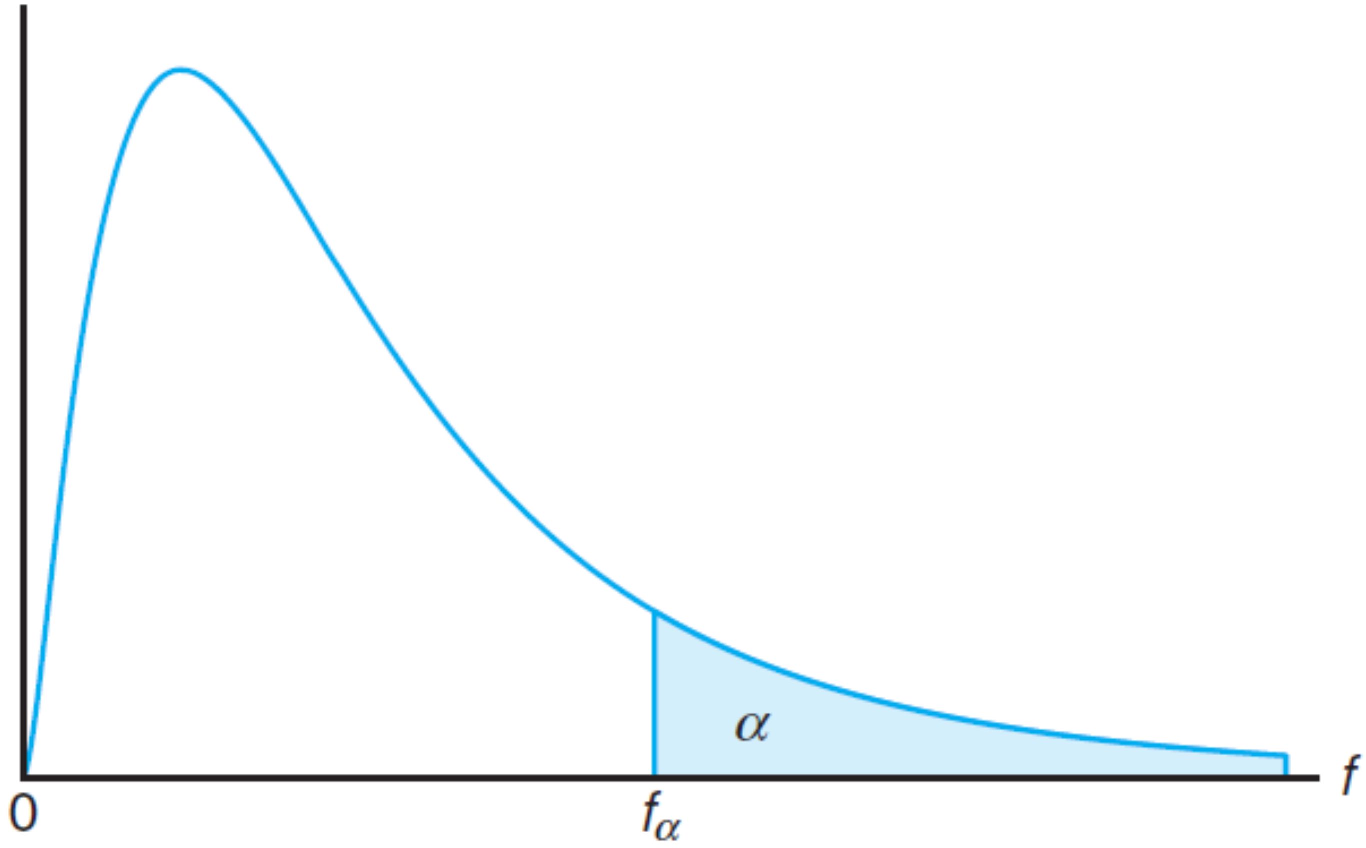


Figure 8.12: Illustration of the f_α for the F-distribution.

Distribusi-F (Catatan)

- Urutan penyebutan degree of freedom cukup penting.
- Kurva distribusi-F dengan degree of freedom v_1 dan v_2 TIDAK SAMA dengan Kurva distribusi-F dengan degree of freedom v_2 dan v_1

- Ingat degree of freedom yang pertama disebut asalnya dari Pembilang pada Rasio yang mendefinisikan F.
- Sedang degree of freedom yang kedua disebut asalnya dari Penyebut.
- Jadi $f_\alpha(v_1,v_2) \neq f_\alpha(v_2,v_1)$

Distribusi-F (Teorema)

- Bila dituliskan $f_\alpha(v_1,v_2)$ untuk nilai f_α dengan v_1 dan v_2 degree of freedom, maka bisa diperoleh

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}.$$

- Contoh: Tentukan nilai f pada kurva distribusi F dengan $v_1=10$ dan $v_2=6$ degree of freedom yang akan menghasilkan area di bawah kurva di sebelah kanan f seluas sebesar $\alpha=0,05$
- Yang ditanya adalah $f_{0,05}(10,6)$
- Dari tabel berikut ditemukan bahwa $f_{0,05}(10,6)=4,06$

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84

Distribusi-F (Contoh)

- Bila yang diminta adalah nilai f pada kurva distribusi F dengan $v_1=6$ dan $v_2=10$ degree of freedom yang akan menghasilkan area di bawah kurva di sebelah kanan f seluas sebesar $\alpha=0{,}95$.

- Maka yang ditanya adalah: $f_{0,95}(6,10)$
- Kita masih bisa menggunakan tabel F dengan $\alpha=0{,}05$ karena (cek Teorema di atas):

$$f_{0,95}(6,10) = \frac{1}{f_{0,05}(10,6)} = \frac{1}{4,06} = 0,246.$$

Distribusi F untuk dua Sampel Variance

- Jika random sampel berukuran n_1 diambil dari populasi berdistribusi normal dengan standar deviasi σ_1.
- Dan random sampel berukuran n_2 diambil dari populasi berdistribusi normal dengan standar deviasi σ_2.
- Maka telah ditemukan bahwa peubah acak:

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \qquad \chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2}$$

berdistribusi chi-squared berturut-turut dengan n_1-1 dan n_2-1 degree of freedom.

- Karena kedua sampel di atas diambil secara random dan terpisah maka χ_1^2 dan χ_2^2 saling independent.

Distribusi F untuk dua Sampel Variance

- Dari uraian di atas bisa dirumuskan teorema berikut.
- Jika S_1 adalah sample standard deviation dari sampel acak beurukuran n_1 yang diambil dari populasi berdistribusi normal dengan standar deviasi σ_1,
- Dan S_2 adalah sample standard deviation dari sampel acak beurukuran n_2 yang diambil dari populasi berdistribusi normal dengan standar deviasi σ_2.
- Maka:

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

memiliki distribusi F dengan $v_1=n_1-1$ dan $v_2=n_2-1$ degree of freedom.

Contoh (PR)

Rerata jumlah air yang diperlukan oleh seorang pada saat melakukan aktifitas fisik adalah 2L dengan standard deviasi 0.7L. Suatu hari akan diadakan wisata alam untuk 50 orang dan disediakan 110L air. Berapakah probabiltas bahwa akan kehabisan air?