Teori Estimasi

Buku Walpole Bab 9

Content

- Statistical Inference
- Single sample: Estimating the Mean
- Two samples: Estimating the Mean
- Single sample: Estimating the Proportion
- Two samples: Estimating the Proportion
- Single sample: Estimating the Variance
- Two samples: Estimating the Variance

Pengantar

- Populasi (Population): Keseluruhan dari observasi yang menjadi perhatian kita => Jumlahnya bisa sangat besar seolah tak berhingga.
- Sampel (Sample) => subset dari populasi:
  - Biasanya dihasilkan dari proses sampling terhadap suatu populasi.
  - Statistical Inference: Bagaimana mendapatkan suatu kesimpulan mengenai suatu populasi berdasarkan sampel.
- Area Statistical Inference:
  - Estimasi
  - Uji Hipotesis: Dalam beberapa bidang => basis dari proses Deteksi.

Point Estimate

- Review: Random Variable biasa dilambangkan dengan huruf besar. Nilai dari random variable dengan huruf kecil.
- Semisal $P(X = x)$, X adalah random variable dan x adalah nilainya.

- Point Estimate dari suatu parameter θ dari suatu populasi adalah suatu nilai $\hat{\theta}$ dari suatu random variable $\widehat{\Theta}$.
- Nilai ini dihitung dari observasi-observasi yang terkandung dalam sebuah sampel yang berperan sebagai estimasi bagi parameter target θ.
- $\widehat{\Theta}$ dinamakan dengan Statistic atau Estimator!

Use Calculation from Sample to Estimate Population Parameter

Population

(select)

Sample

(calculate)

(describes)

Parameter

(to estimate)

Point Estimate

$$p = ?$$

$$\hat{p} = 63\%$$

Example

- Consider a population of the entire women in India.
- We are interested in the actual mean of the number of children born to each woman.

- However, figuring out the actual mean of the number of children per woman would be impractical
- How could one estimate the parameter of interest in this situation?

Example

Solution

- An intuitively appealing estimate of a population mean, μ, is the sample mean, $\bar{x}$, computed from a random sample of n observations from the target population.
- Assume, for example, that we obtain a random sample of size $n = 30$ women whose number of children is recorded.
- We then compute the value of the sample mean to be $\bar{x} = 3.05$ children.
- Here, $\bar{x}$ provides a point estimate of the population mean μ.

Contoh

- Tiga buah Random variable X_0, X_1, dan X_2 memiliki nilai yang diberikan oleh sampel-sampel dari suatu populasi (secara berturut2) yaitu $x_0 = 2, x_1 = 5$, dan $x_2 = 11$.
- Mean aktual dari populasi (diasumsikan tidak diketahui) $\mu = 4$

Contoh

- Beberapa kemungkinan point estimator untuk mean
  - Sample Mean: $\bar{X}$ yang nilainya diberikan oleh $\bar{x}$=$(x_0 + x_1 + x_2)/3 = 6$
  - Sample Median: $\tilde{X}$ yang nilainya diberikan oleh nilai tengah yaitu $x_1 = 5$
  - Sampel yang pertama yaitu: $x_0 = 2$

- Mana yang terbaik? Untuk kasus ini adalah sample median => **bisa berbeda untuk sampel yang berbeda** (misalnya 1, 3, dan 7 atau 4, 7, dan 11).

Estimator Terbaik

- Jika suatu random variable $\widehat{\Theta}$ adalah suatu point estimator untuk parameter θ dan nilai $\widehat{\Theta}$ diberikan oleh $\hat{\theta}$ maka diinginkan $\widehat{\Theta}$ memiliki sifat unbiased, y.i yang memenuhi:

$$\boxed{\mu_{\widehat{\Theta}} = E\left(\widehat{\Theta}\right) = \theta}$$

- Ketiga point estimator pada contoh di slide sebelumnya semuanya adalah unbiased estimator! Mengapa?
- Di samping itu diinginkan juga agar **variance** dari estimator $\widehat{\Theta}$ **sekecil mungkin**

Efficient Estimator

- Jika $\widehat{\Theta}_1$ dan $\widehat{\Theta}_2$ adalah unbiased estimator dari parameter yang sama yaitu θ maka estimator yang diinginkan adalah estimator yang distribusinya memiliki variance yang lebih kecil.
- Jika: $\sigma_1^2 < \sigma_2^2$ maka $\widehat{\Theta}_1$ adalah estimator yang **lebih efisien** daripada $\widehat{\Theta}_2$.

- Jika dipertimbangkan seluruh unbiased estimator untuk parameter θ, estimator dengan variance terkecil dikenal sebagai estimator terefisien.

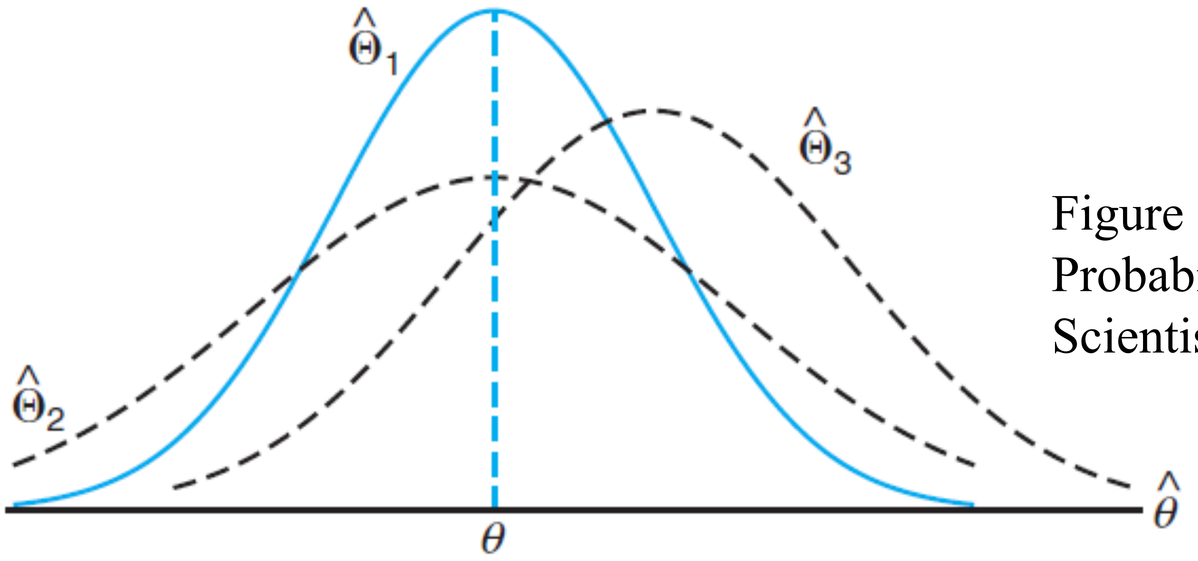


Figure 9.1 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

Interval Estimate

- Sebuah point estimate yang diperoleh dari sample tidak bisa diharapkan akan persis sama dengan parameter populasi yang diestimasi.
- Interval Estimate dari suatu parameter: Sebuah interval yang di dalamnya kita berharap bahwa **true value** dari parameter populasi yang diestimasi berada di dalam interval tersebut.

- An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$ where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\widehat{\Theta}$ (which is $\hat{\theta}$) for a particular sample and the distribution of $\widehat{\Theta}$.

Interval Estimate

- Contoh: Untuk tinggi badan mahasiswi UGM, nilai mean yang sesungguhnya tidak diketahui.
  - Contoh Point Estimate untuk mean: $\hat{\theta} = 160$cm (kita berharap mean yang sesungguhnya juga 160cm)
  - Contoh Interval Estimate:

$$157 < \mu < 163$$

Confidence Interval

- Katakanlah ditemukan nilai $\widehat{\Theta}_L$ and $\widehat{\Theta}_U$ sedemikian rupa sehingga

$$P\left(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U\right) = 1 - \alpha$$

di mana $0 \leq \alpha \leq 1$.

  - Maka interval $\widehat{\Theta}_L < \theta < \widehat{\Theta}_U$ dikatakan sebagai 100(1-α)% confidence interval.
  - Nilai (1-α) dikatakan sebagai confidence coefficient.
  - Semakin lebar confidence interval, kita semakin yakin bahwa interval tersebut mengandung nilai parameter yang diestimasi.
  - Mana yang lebih baik: 95% yakin bahwa rerata usia transistor adalah antara 6 dan 7 tahun atau 99% yakin bahwa rerata usia transistor adalah antara 3 dan 10 tahun?

Estimating the Mean: Single Sample (σ known)

- Asumsikan sebuah populasi akan dicari true-mean-nya (μ).
- Asumsikan standar deviasi populasi yaitu σ → diketahui.
- Diketahui beberapa sampel $x_1, x_2, \ldots, x_N$. Point Estimator yang dipilih adalah sample mean $\bar{X}$ yang nilainya diberikan oleh:

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

  - $\bar{X}$ sendiri adalah random variable
  - Motivasi: $\bar{X}$ adalah unbiased estimator. Variance dari $\bar{X}$ akan turun seiring bertambahnya n.

$$\mathrm{Var}(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma^2}{n}$$

Estimating the Mean: Single Sample (σ known)

- Apabila jumlah sampel *n* cukup banyak → menurut Central Limit Theorem, distribusi dari sample mean $\bar{X}$ dapat didekati dengan distribusi Normal/Gaussian, irrelevant dengan distribusi dari populasi itu sendiri.
- Mean dari sample mean $\bar{X}$ adalah mean dari populasi (unbiased estimator), y.i:

$$E[\bar{X}] = \mu_{\bar{X}} = \mu$$

- Gunakan fakta ini untuk temukan interval estimate → gunakan bantuan kurva distribusi normal/Gaussian standard (mean=0, variance=1)

Estimating the Mean: Single Sample (σ known)

- Temukan 100(1-α)% confidence interval dengan kurva distribusi normal standard.

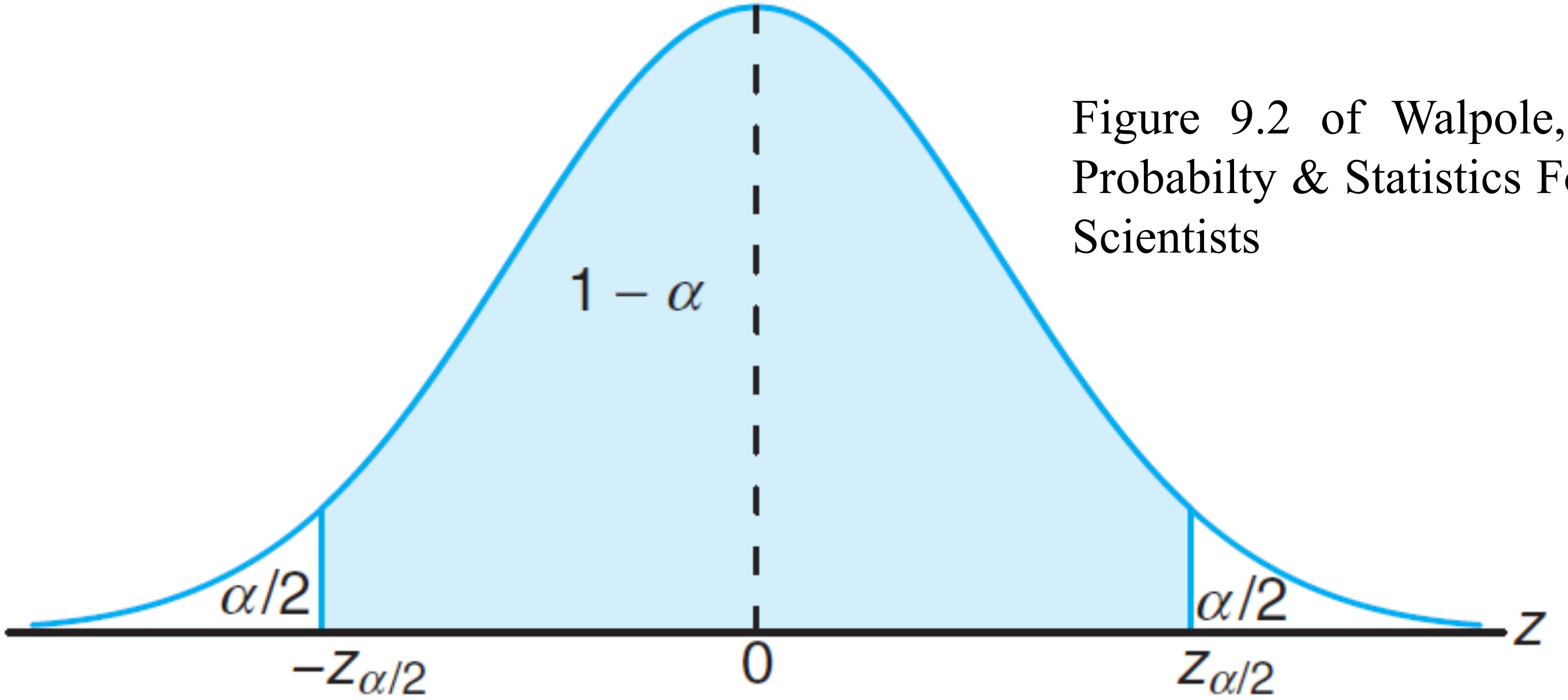


Figure 9.2 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

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

Estimating the Mean: Single Sample (σ known)

- Didapatkan:

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

- Jika $\bar{x}$ adalah mean dari sampel random berukuran n dari sebuah populasi yang variance-nya (σ^2) diketahui, maka 100(1-α)% confidence interval untuk μ diberikan oleh:

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

dengan $z_{\alpha/2}$ adalah nilai z sedemikan rupa sehingga area di bawah kurva distribusi normal standar mulai dari z ke kanan memiliki luas $\alpha/2$.

Example 9.2 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

Rerata konsentrasi seng yang diperoleh dari sampel pengukuran di 36 lokasi berbeda pada suatu sungai adalah 2.6 gr/ml. Temukan 95% dan 99% confidence interval untuk true mean dari konsentrasi seng pada sungai. Standar dev. Populasi diketahui 0.3gr/ml.

Jawab: Point Estimate $\bar{x} = 2.6$

Untuk $100(1-\alpha)\% = 95\%$ => diperoleh dari tabel distribusi normal standard: $z_{0.025} = 1.96$ => 95%-confidence interval adalah:

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right) \quad \text{atau} \quad 2.50 < \mu < 2.70$$

Untuk 99% => diperoleh dari tabel distribusi normal standard: $z_{0.005} = 2.575$ => 99%-confidence interval adalah:

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right) \quad \text{atau} \quad 2.47 < \mu < 2.73$$

Estimating the Mean: Single Sample (σ known) – One Sided Confidence Bound

Lower one-sided bound:

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

Upper one-sided bound:

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

- Jika $\bar{x}$ adalah mean dari sampel random berukuran n dari sebuah populasi yang variance-nya (σ^2) diketahui, maka one-sided 100(1-α)% confidence bound untuk μ diberikan oleh:

Upper one-sided bound: $\bar{x} + z_\alpha \sigma/\sqrt{n};$

Lower one-sided bound: $\bar{x} - z_\alpha \sigma/\sqrt{n}.$

Example 9.4 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

- Dalam suatu tes psikologis, 25 subyek dipilih acak dan waktu reaksi mereka terhadap stimulus tertentu diukur. Variansi waktu reaksi diketahui dari pengalaman sebelumnya sebesar 4sec^2 dan distribusi waktu reaksi mendekati Gaussian/normal. Waktu rata-rata yang diperoleh dari 25 subyek adalah 6.2 sec. Temukan upper 95% bound untuk mean waktu reaksi.

- Jawab: Upper 95% bound diberikan sbb:

$$\bar{x} + z_\alpha \sigma/\sqrt{n} = 6.2 + (1.645)\sqrt{4/25} = 6.2 + 0.658$$
$$= 6.858 \text{ seconds.}$$

- Jadi kita yakin 95% bahwa mean waktu reaksi kurang dari 6.858 sec.

Estimating the Mean: Single Sample (σ unknown)

- Bagaimana jika variance yang sesungguhnya (σ^2) dari populasi tidak diketahui?
- Kita bisa gunakan sample variance S^2:

$$s^2 = \frac{1}{n-1}\sum_{n=1}^{n}(x_n - \bar{x})^2$$

Dimana $\bar{x}$ adalah nilai dari sample mean dan N ukuran sampel.

- Sample standard deviation S adalah random variable dan unbiased estimate bagi σ, jadi: $E[S] = \sigma$.
- Jika sampel diambil dari populasi yang memiliki DISTRIBUSI NORMAL (penting!!), maka random variable

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

mematuhi Student t-distribution dengan $N - 1$ degrees of freedom, di mana μ adalah true mean dan $\bar{X}$ adalah random variable untuk sample mean.

t-distribution

- The t-distribution is very much like the z-distribution (which is Gaussian).
- In particular, both are symmetric, bell-shaped, and have a mean of 0.
- However, the distribution of t depends on a quantity called its degrees of freedom (df), which is equal to $(n - 1)$ when estimating a population mean based on a small sample of size n.

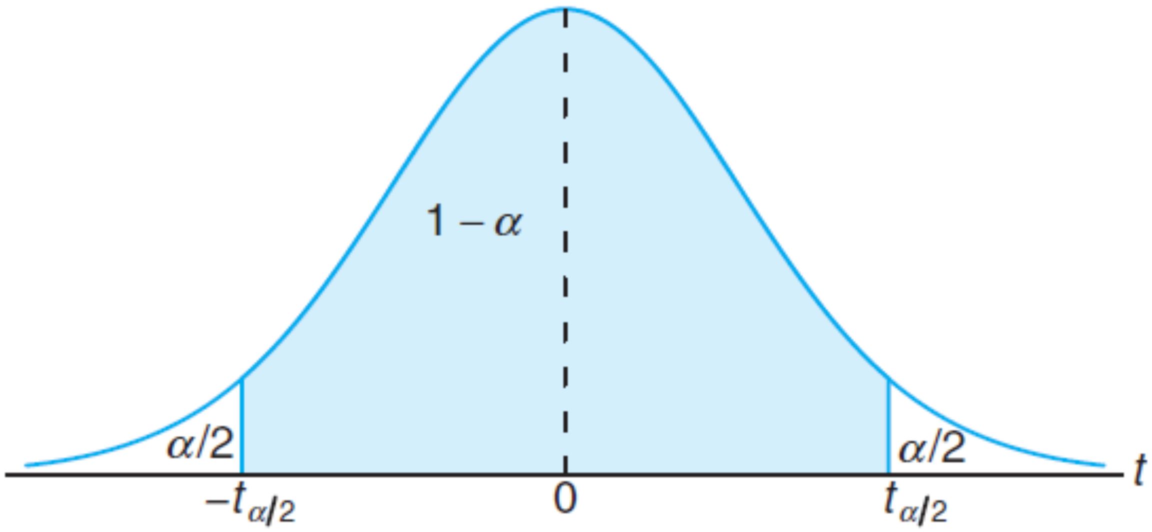


Figure 9.5 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists: Kurva Student-*t* distribution.

Random variable T di atas bisa dipakai untuk menentukan confidence interval: menggunakan S sebagai pengganti σ:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $t_{\alpha/2}$ is the t-value with $N-1$ degree of freedom, above which we find an area of $\alpha/2$.

- Bentuk kurva simetris. Substitusikan nilai T ke dalam formula probabilitas:

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

Estimating the Mean: Single Sample (σ unknown)

- Jika $\bar{x}$ dan s adalah nilai mean dan standard deviasi dari suatu sampel random yang diambil dari populasi berdistribusi normal dengan unknown variance σ^2, maka $100(1-\alpha)\%$ confidence interval untuk μ diberikan oleh:

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

di mana $t_{\alpha/2}$ adalah t-value dengan $v = N - 1$ degrees of freedom, dan luas area di bawah kurva mulai dari titik t tersebut adalah $\alpha/2$.

- Bila yang diinginkan adalah one-sided confident bounds maka upper dan lower $100(1-\alpha)\%$ bounds diberikan oleh:

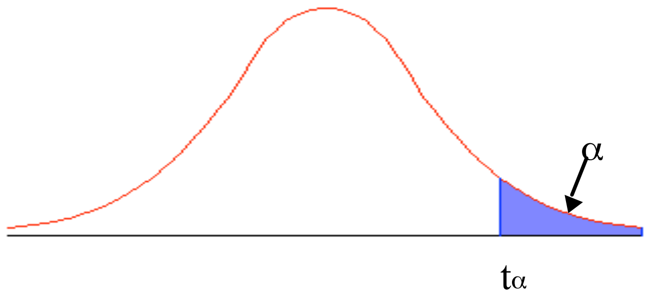
$$\bar{x} + t_{\alpha}\frac{s}{\sqrt{n}} \quad \text{dan} \quad \bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}.$$

di mana t_{α} adalah t-value dengan $v = N - 1$ degrees of freedom, dan luas area di bawah kurva mulai dari titik t tersebut adalah α.

Some values for Student’s t-distribution

Table

Some values for Student's t-distribution



Degrees of freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.102	3.852	4.221
14	1.345	1.760	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073

Contoh

Tujuh bejana diambil dari populasi bejana2 berisikan asam sulfida. Isi asam sulfida di ketujuh bejana tsb. adalah 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, dan 9.6 liter. Temukan 95% confidence interval untuk mean dari isi populasi bejana2 jika distribusi dari isi bejana2 tersebut bisa didekati dengan distribusi normal.

Jawab: Point Estimate dari mean adalah sample mean: $\bar{x}$=10.

- Sample standard deviation: $s = 0.283$ (bagaimana menghitungnya).
- Dari tabel distribusi Student-*t* diperoleh $t_{0.025}$=2.447 untuk $v = 6$ degrees freedom.
- Jadi 95% confidence interval untuk μ:

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right) \Longrightarrow 9.74 < \mu < 10.26.$$

Konsep Confidence Interval untuk Sampel Berukuran Besar

- Pada saat ukuran sampel n cukup besar → $n \geq 30$, dan meski distribusi dari populasi tidak bisa dianggap normal (Gaussian): s dapat digunakan untuk menggantikan σ dan confidence interval

  $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ bisa dipakai sebagai aproximasi.

- Sering dikenal dengan: Konsep Large-Sample Confidence Interval.

- Contoh: Dari populasi mahasiswa UGM, dikoleksi sampel nilai test TPA dari 500 mahasiswa. Diperoleh sample mean 501 dan sample standard deviation 112. Temukan 99% confidence interval pada mean nilai test TPA dari seluruh mahasiswa UGM.

- Jawab: Ukuran sampel cukup besar => aproksimasi distribusi normal bisa dipakai. Dari tabel distribusi normal didapat $z_{0.005}$=2.575. 99% confidence interval untuk μ:

$$501 \pm (2.575)\left(\frac{112}{\sqrt{500}}\right) = 501 \pm 12.9$$

Bahan Bacaan Mandiri

- Probability and Statistics For Engineers and Scientists (Walpole, Myers, Myers & Ye, 9th edition)
  - Section 9.5: Standard Error of a Point Estimate
  - Section 9.6: Prediction Intervals
  - Section 9.7: Tolerance Limits
  - Section 9.9: Paired Observations
- Tetap jadi bahan materi untuk ujian.

Estimation of the difference between two population means: Known Variance

- Assume: We have 2 populations (independent from each other) with mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. From each population, we obtain sample of size n_1 and n_2, respectively.
- Point estimator untuk $\mu_1 - \mu_2$ adalah $\bar{X}_1 - \bar{X}_2$ dimana $\bar{X}_k$ adalah sample mean untuk populasi ke-k yang nilainya diberikan oleh $\bar{x}_k$.
  - Untuk sampel yang berukuran cukup besar ($n_k > 29$), $\bar{X}_k$ bisa didekati dengan distribusi Gaussian/normal (Central Limit Theorem).
  - Akibatnya, random variable $\bar{X}_1 - \bar{X}_2$ bisa pula didekati dengan distribusi normal dengan mean dan variance:

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

$$\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Estimation of the difference between two population means: Known Variance

- Pencarian interval estimator: memanfaatkan kenormalan distribusi $\bar{X}_1 - \bar{X}_2$.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

- Untuk temukan $100(1-\alpha)\%$ confidence interval bagi $\mu_1 - \mu_2$:

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

$$P\left(-z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < z_{\alpha/2}\right) = 1 - \alpha$$

Estimation of the difference between two population means: Known Variance

- Jika $\bar{x}_1$ dan $\bar{x}_2$ adalah sample means dari 2 sampel random independent yg masing2 berukuran n_1 and n_2 dari 2 buah populasi yang variance-nya (σ_1^2 and σ_2^2) diketahui, maka 100(1-α)% confidence interval untuk $\mu_1 - \mu_2$ diberikan oleh:

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

dengan $z_{\alpha/2}$ adalah nilai *z* sedemikan rupa sehingga area di bawah kurva distribusi normal standar mulai dari *z* ke kanan memiliki luas $\alpha/2$.

Contoh

- Akan dievaluasi pemakaian bensin mesin A dan mesin B dalam ukuran km/galon. 50 eksperimen yang independent dilaksanakan untuk mesin A dan 75 eksperimen untuk mesin B.

- Dari eksperimen2 di atas diperoleh sample mean pemakaian bensin untuk mesin A adalah 36 km/galon dan untuk mesin B adalah 42km/galon.

- Diasumsikan standar deviasi populasi pemakaian bensin untuk mesin A dan B adalah $\sigma_A = 6$ dan $\sigma_B = 8$

- Temukan 96% confidence interval untuk $\mu_B - \mu_A$ di mana μ_A dan μ_B adalah mean populasi pemakaian bensin untuk mesin A dan B.

Contoh (Lanjutan)

- Solusi: Point Estimate:

$$\bar{x}_B - \bar{x}_A = 42 - 36 = 6$$

- (1- α)=0.96 → Confidence Interval: $\alpha = 0.04$
- Dari table standard normal distribution: $z_{0.02}$=2.05
- 96% Confidence Interval:

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

$$3.43 < \mu_B - \mu_A < 8.57$$

Estimating the difference between 2 population means: Unknown but Equal Variance

- Diketahui bahwa $\sigma_1^2 = \sigma_2^2 = \sigma^2$ namun nilainya tidak diketahui.
- Sample diambil dari Populasi yang distribusinya mendekati normal/Gaussian
- Untuk kasus ini bisa digunakan sample standard deviation S_1 dan S_2.
- Gunakan sifat bahwa random variabel

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \quad \text{dan} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2}$$

memiliki distribusi chi-squared dengan $n_1 - 1$ dan $n_2 - 1$ degrees of freedom..

Estimating the difference between 2 population means: Unknown but Equal Variance

- Interval Estimate bisa ditemukan dengan random variabel/statistik *T* yang mematuhi student *t*-distribution. Detail derivasi bisa ditemukan di Section 9.8

- 100(1-α)% confidence interval untuk $\mu_1 - \mu_2$ bisa dicari dengan:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$$

- Di mana

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Estimating the difference between 2 population means: Unknown but Equal Variance

- Dgn demikian, untuk kasus ini, 100(1-α)% confidence interval untuk $\mu_1 - \mu_2$ diberikan oleh:

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

di mana $t_{\alpha/2}$ adalah t-value dengan $v = n_1 + n_2 - 2$ degrees of freedom sedemikian rupa shg luas area di bawah kurva distribusi-*t* mulai dari titik *t* tersebut ke arah kanan adalah $\alpha/2$.

Sedangkan s_p adalah nilai dari random variable S_p terhitung di atas.

Estimating the difference between 2 population means: Unknown and Unequal Variance

- Diketahui $\bar{x}_1$ dan s_1 dan $\bar{x}_2$ dan s_2 adalah sample means dan sample standar deviasi dari 2 sampel random independent yg masing2 berukuran n_1 and n_2 dari 2 buah populasi.
- Kedua populasi tersebut bisa didekati dengan distribusi normal namun variancenya tidak diketahui dan besarnya tidak sama
- Maka 100(1-α)% confidence interval untuk $\mu_1 - \mu_2$ diberikan oleh:

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

Estimating the difference between 2 population means: Unknown and Unequal Variance

Di mana $t_{\alpha/2}$ adalah nilai *t* dengan

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

degrees of freedom, sedemikian rupa shg luas area di bawah kurva distribusi-*t* mulai dari titik *t* tersebut ke arah kanan adalah $\alpha/2$.

Contoh

- Suatu studi dilakukan oleh Departemen Ilmu Hewan suatu PT di Indonesia guna melakukan estimasi jumlah senyawa fosfor pada 2 stasiun yang berbeda di Sungai Brantas. Senyawa fosfor ini diukur dalam mg/liter. 15 observasi dikoleksi dari stasiun 1 dan 12 observasi dikoleksi dari stasiun 2. 15 observasi dari stasiun 1 memiliki rerata kandungan fosfor sebesar 3.84 mg/liter dan standar deviasi 3.07 mg/liter. 12 observasi dari stasiun 2 memiliki rerata kandungan fosfor sebesar 1.49 mg/liter dan standar deviasi 0.8 mg/liter. Temukan 95% confidence interval untuk selisih true mean kandungan fosfor pada 2 stasiun di atas, dengan mengasumsikan bahwa obserasi berasal dari populasi dengan distribusi normal dengan variance yang berbeda.

Contoh: Solusi

- Untuk stasiun 1 diperoleh: $\bar{x}_1 = 3.84,\ s_1 = 3.07,\ n_1 = 15$
  - Untuk stasiun 2 diperoleh: $\bar{x}_2 = 1.49,\ s_2 = 0.80,\ n_2 = 12$
  - Karena populasi variances untuk kedua stasiun diasumsikan berbeda, 95% confidence interval untuk μ_1-μ_2 didekati dengan t-distribution dengan degree of freedom:

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

  - Point estimate untuk μ_1-μ_2:

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

  - Dari tabel distribusi student-t, diperoleh $t_{0.025}$=2.120 untuk v=16 degrees of freedom.

Contoh: Solusi

- 95% confidence interval untuk μ_1-μ_2 adalah

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

$$0.60 < \mu_1 - \mu_2 < 4.10$$

  - Kita 95% yakin bahwa interval 0.6 $<\mu_1$-$\mu_2<$4.1 mg/liter memuat selisih dari true average dari kandungan fosfor pada 2 lokasi di atas.

Single Sample: Estimasi Proporsi

- Pada eksperimen Binomial, point estimator untuk proporsi p diberikan oleh statistik:

$$\hat{P} = \frac{X}{n}$$

  dengan X adalah jumlah sukses per n percobaan. Nilai estimate dari statistik di atas diberikan oleh:

$$\hat{p} = \frac{x}{n}$$

- Jika p tidak terlalu dekat dengan 0 atau 1, confidence interval untuk p bisa diformulasikan dengan mempertimbangkan distribusi $\hat{P}$.
- Indikasikan kegagalan dalam tiap percobaan binomial dengan nilai 0 dan sukses dengan nilai 1 → x: jumlahan n variabel masing2 bernilai 0 atau 1

Single Sample: Estimasi Proporsi

- $\hat{p}$ => sample mean dari n variabel.
- Central Limit Theorem: Untuk n cukup besar => $\hat{P}$ bisa didekati dengan distribusi Gaussian dengan mean:

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

dan variance

$$\sigma^2_{\widehat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

Dengan demikian, 100% (1-α) confidence interval diberikan oleh:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \text{ with } Z = \frac{\widehat{P} - p}{\sqrt{pq/n}}$$

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

Single Sample: Estimasi Proporsi

- Untuk harga n besar, nilai p dan q di dalam operasi akar kuadrat bisa didekati dengan $\hat{p} = x/n$ dan $\hat{q} = 1 - \hat{p}$.

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

- Atau jika ingin disolusikan nilai p dari pertidaksamaan kuadrat di atas, y.i.

$$-z_{\alpha/2} < \frac{\widehat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2}$$

Diperoleh confidence interval untuk p dengan batasan

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}} \longrightarrow \textbf{Metode II}$$

Single Sample: Estimasi Proporsi

Large-Sample Confidence Interval for p: Jika $\hat{p}$: proporsi sukses pada random sampel berukuran n dan $\hat{q} = 1 - \hat{p}$, aproksimasi dari 100(1-α)% confidence interval untuk parameter binomial p diberikan oleh

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

atau

Metode II:

$$\frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n}}{1 + \frac{z^2_{\alpha/2}}{n}} - \frac{z_{\alpha/2}}{1 + \frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}} < p < \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n}}{1 + \frac{z^2_{\alpha/2}}{n}} + \frac{z_{\alpha/2}}{1 + \frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}}$$

Single Sample: Estimasi Proporsi

- Catatan:
  - Saat n kecil dan p (yang tak diketahui) bernilai dekat dengan 0 atau 1: formula confidence interval di atas barangkali tidak akurat.
  - Biasanya diperlukan baik $n\hat{p} \geq 5$ dan $n\hat{q} \geq 5$
  - Metode 2 lebih kompleks dan kelebihan metode 2 dalam hal akurasi dibanding metode 1 tidak terlalu besar saat ukuran sampel cukup besar.
  - Metode 1 lebih umum dipakai.

Single Sample: Estimasi Proporsi (Contoh)

- Dari sebuah sampel random n=500 keluarga di suatu kota yang memiliki perangkat televisi, ada x=340 keluarga yang berlangganan HBO. Temukan 95% confidence interval bagi proporsi yang sebenarnya dari seluruh keluarga dengan pesawat televisi di kota tersebut yang berlangganan HBO.

- Jawab: Point estimate dari p: $\hat{p} = 340{=}500 = 0{:}68{:}$
- Dari tabel distribusi Gaussian: $z_{0:025} = 1{:}96$
- Dengan menggunakan metode 1, pada slide sebelumnya, 95% confidence interval untuk p adalah:

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

$$0.6391 < p < 0.7209$$

Single Sample: Estimasi Proporsi (Contoh)

- Jika memakai metode 2 diperoleh

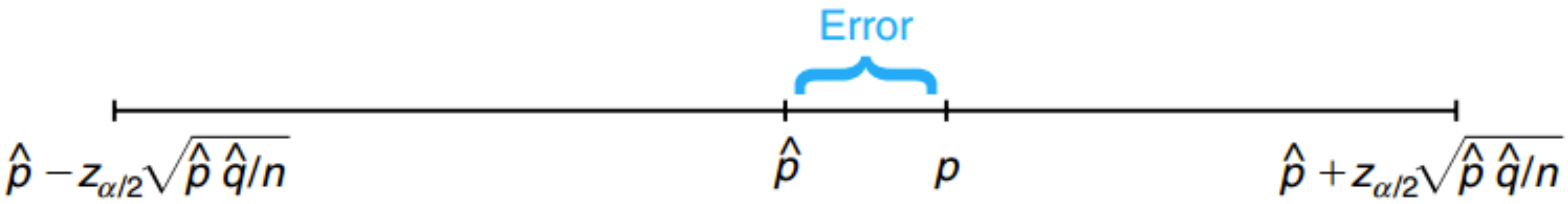
$$\frac{0.68 + \frac{1.96^2}{(2)(500)}}{1 + \frac{1.96^2}{500}} \pm \frac{1.96}{1 + \frac{1.96^2}{500}} \sqrt{\frac{(0.68)(0.32)}{500} + \frac{1.96^2}{(4)(500^2)}} = 0.6786 \pm 0.0408$$

sehingga: $0.6378 < p < 0.7194$

Tampak bahwa hasil tidak jauh berbeda untuk kedua metode dengan *n* cukup besar (500).

Single Sample: Estimasi Proporsi

- Jika $\hat{p}$ adalah estimate dari p, kita dapat 100(1-α)% yakin bahwa error yang dihasilkan tidak akan melebihi $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.



- Jika $\hat{p}$ adalah estimate dari p, kita dapat 100(1-α)% yakin bahwa error yang dihasilkan tidak akan melebihi nilai error sebesar e saat ukuran sampel bernilai:

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

Diperoleh dengan faktor batasan pada convergence interval:

$$z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} = e.$$

Agak misleading

Single Sample: Estimasi Proporsi

- Pernyataan terakhir agak misleading sebab disebutkan di situ bahwa kita harus gunakan $\hat{p}$ untuk tentukan ukuran sample n padahal $\hat{p}$ diperoleh dari sampel.
- Bisa dilakukan, kita ambil sampel awal dengan ukuran sampel di atas 30 untuk menentukan $\hat{p}$ lalu kita gunakan $\hat{p}$ yang diperoleh untuk menentukan ukuran sampel baru yang diperlukan untuk memperoleh tingkat akurasi tertentu.
- Ingat: Jumlah sampel selalu bilangan bulat positif.

Contoh

- Contoh: Tentukan ukuran sampel yang diperlukan untuk membuat kita 95% yakin bahwa estimasi untuk *p* pada contoh sebelumnya adalah dalam kisaran error sebesar 0.02 dari nilai yang sebenarnya.

- Jika kita perlakukan 500 keluarga pada contoh sebelumnya sebagai sampel awal, yang lalu menghasilkan: $\hat{p}$=0.68 maka bisa kita gunakan nilai $\hat{p}$ untuk cari nilai n yang memang diperlukan untuk mencapai e=0.02:

$$n = \frac{(1.96)^2(0.68)(0.32)}{(0.02)^2} = 2089.8 \approx 2090.$$

Single Sample: Estimasi Proporsi

- Jika $\hat{p}$ adalah estimate dari p, kita dapat **setidaknya** 100(1-α)% yakin bahwa error yang dihasilkan tidak akan melebihi nilai error sebesar e saat ukuran sampel bernilai:

$$n = \frac{z_{\alpha/2}^2}{4e^2}$$

- Hal ini bermanfaat saat kita tidak punya nilai estimasi awal tentang $\hat{p}$ untuk menentukan jumlah sampel n. Di sini kita gunakan fakta bahwa:

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}$$

Bernilai maksimal saat $\hat{p}\hat{q} = \hat{p}(1-\hat{p})$ maksimal yaitu saat $\hat{p}$=0.5.

Two Sample: Estimasi Selisih Dua Proporsi

- Diketahui 2 buah populasi binomial dengan parameter binomial p_1 dan p_2.
- Contoh : Proporsi penderita kanker paru2 pada populasi perokok dan proporsi penderita kanker paru2 pada populasi bukan perokok.
- Akan dicari selisih antara dua parameter binomial di atas (atau pada contoh: selisih proporsi penderita kanker paru2 pada populasi perokok dan bukan perokok), y.i: p_1 - p_2
- Point Estimator: $\hat{P}_1$- $\hat{P}_2$ => Nilai point estimate yang dipakai $\hat{p}_1$- $\hat{p}_2$ dengan $\hat{p}_1 = x_1/n_1$ dan $\hat{p}_2 = \frac{x_2}{n_2}$ di mana x_1 dan x_2 berturut-turut adalah jumlah penderita kanker paru2 pada populasi perokok dan bukan perokok dari sampel yang diambil (berukuran n_1 untuk sampel perokok dan n_2 untuk bukan perokok).

Two Sample: Estimasi Selisih Dua Proporsi

- Dari kasus Single sample: $\hat{P}_1$ dan $\hat{P}_2$ masing2 bisa didekati dengan distribusi normal saat n_1 dan n_2 cukup besar.
- Random variable $\hat{P}_1$- $\hat{P}_2$ bisa didekati dengan distribusi normal dengan mean:

$$\mu_{\hat{P}_1 - \hat{P}_2} = p_1 - p_2$$

- Dan variance (lihat kasus Satu Sample):

$$\sigma^2_{\hat{P}_1 - \hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

- Catatan: Sampel pada populasi 1 independent terhadap sampel pada populasi 2.
- Dengan demikian, $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- di mana:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{p_1 q_1 / n_1 + p_2 q_2 / n_2}}$$

Two Sample: Estimasi Selisih Dua Proporsi

$$P\left[-z_{\alpha/2} < \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}} < z_{\alpha/2}\right] = 1 - \alpha.$$

Seperti kasus single sample, nilai p_1, p_2, q_1, dan q_2 di dalam operasi akar di atas bisa digantikan oleh point estimatenya yaitu $\hat{p}_1$, $\hat{p}_2$, $\hat{q}_1$=1- $\hat{p}_1$, dan $\hat{q}_2$=1- $\hat{p}_2$, jika

$$n_1\hat{p}_1,\ n_1\hat{q}_1,\ n_2\hat{p}_2, n_2\hat{q}_2 \geq 5$$

Kesimpulan bagi Large-Sample Confidence Interval untuk p_1-p_2:

Jika $\hat{p}_1$ dan $\hat{p}_2$ berturut-turut adalah proporsi sukses di sampel berukuran n_1 dan n_2, dan $\hat{q}_1$=1-$\hat{p}_1$ dan $\hat{q}_2$=1-$\hat{p}_2$, maka aproksimasi 100(1-α)% confidence interval bagi selisih dua parameter binomial p_1-p_2 diberikan oleh:

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

Contoh

- Sedang dipertimbangkan suatu perubahan dalam proses manufaktur di suatu perusahaan. Sampel diambil dari hasil proses manufaktur yang lama dan dari hasil proses manufaktur yang baru untuk mengecek apakah proses yang baru mengindikasikan perbaikan. Jika 75 dari 1500 produk dari proses yang lama ditemukan rusak dan 80 dari 2000 produk dari proses yang baru ditemukan rusak, temukan 90% confidence interval untuk nilai selisih yang sesungguhnya dari proporsi kerusakan barang pada proses yang lama dan proses baru.

Contoh

- Solusi: Dicari p_1 dan p_2 berturut-turut adalah proporsi produk rusak pada proses manufaktur lama dan baru.
- Dengan demikian $\hat{p}_1$=75/1500=0.05 dan $\hat{p}_2$=80/2000=0.04.
- Point estimate untuk p_1 - p_2 adalah $\hat{p}_1$-$\hat{p}_2$=0.01
- Dari tabel distribusi normal diperoleh $z_{0.05}$=1.645. Dengan demikian:

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

- 90% confidence interval yang dicari:

$$-0.0017 < p_1 - p_2 < 0.0217$$

Single Sample: Estimasi Variance

- Katakanlah sampel berukuran n diambil dari populasi berdistribusi normal dengan variance σ^2 dan lalu sample variance s^2 dihitung (yang merupakan nilai dari random variabel S^2).
- $s^2 \rightarrow$ point estimate untuk σ^2 dan S^2 disebut estimator untuk σ^2.
- Statistik/Random variable berikut:

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

memenuhi distribusi chi-squared dengan n-1 degrees of freedom jika sampel diambil dari populasi berdistribusi normal. Bisa dituliskan:

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

Dengan $\chi^2_{1-\alpha/2}$ dan $\chi^2_{2-\alpha/2}$ berturut-turut adalah nilai Chi-squared dengan $n-1$ degrees of freedom yang menyisakan area sebesar 1-α/2 dan α/2 di sebelah kanan.

Single Sample: Estimasi Variance

Diperoleh:

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

Setelah beberapa operasi matematika diperoleh:

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

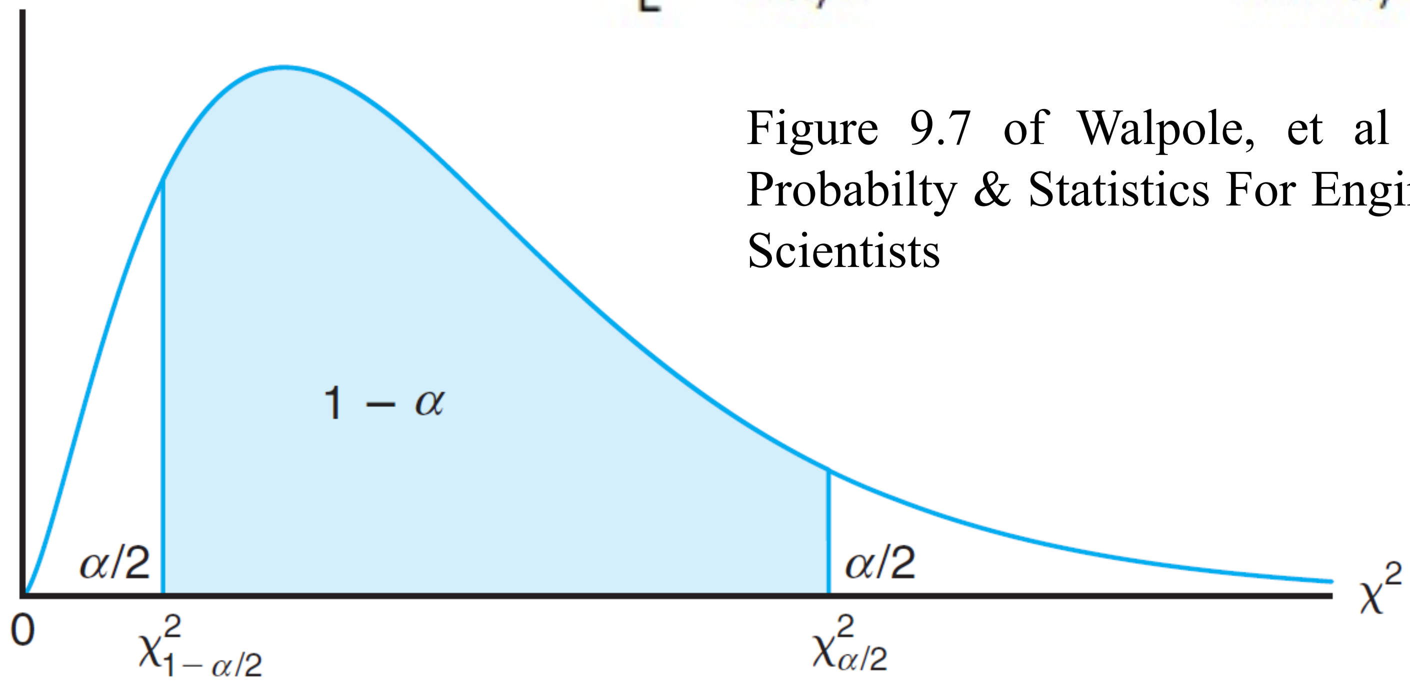


Figure 9.7 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

Single Sample: Estimasi Variance

Kesimpulan: Jika s^2 adalah sample variance dari sampel random berukuran n dari populasi berdistribusi normal, 100(1-α)% confidence interval untuk σ^2 adalah

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

Dengan $\chi^2_{1-\alpha/2}$ dan $\chi^2_{2-\alpha/2}$ berturut-turut adalah nilai Chi-squared dengan $v = n - 1$ degrees of freedom yang menyisakan area sebesar 1-α/2 dan α/2 di sebelah kanan.

Contoh

- Berikut adalah data berat badan dari 10 anak dalam kg di suatu sekolah dasar: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, dan 46.0. Temukan 95% confidence interval untuk variance berat badan dari seluruh anak di SD jika diasumsikan populasi berat badan tsb memenuhi distribusi normal

- Solusi:

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

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

Contoh

- Untuk α=0.05 (95%-confidence interval) diperoleh dari Tabel untuk Chi-Square distribution dengan v=9 degrees of freedom:

$$\chi^2_{0.025} = 19.023 \text{ dan } \chi^2_{0.975} = 2.700$$

- 95%-confidence interval untuk σ^2 adalah:

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700}$$

atau: $0.135 < \sigma^2 < 0.953$

Two Samples: Estimasi Rasio Dua Variances

- Rasio variance dua populasi: $\sigma_1^2 = \sigma_2^2$
  - Point estimate untuk $\sigma_1^2 = \sigma_2^2$ diberikan oleh rasio sample variance: $s_1^2 = s_2^2$
  - Statistik $S_1^2 = S_2^2$ $\Rightarrow$ Estimator bagi $\sigma_1^2 = \sigma_2^2$
  - Jika σ_1 dan σ_2 adalah standar deviasi dari populasi berdistribusi normal maka interval estimate untuk $\sigma_1{}^2 / \sigma_2{}^2$ bisa digunakan statistik:

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

di mana random variable F memiliki distribusi F dengan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$ degrees of freedom.

Two Samples: Estimasi Rasio Dua Variances

- Sehingga bisa dituliskan

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

- Di mana $f_{1\text{-}\alpha/2}(v_1,v_2)$ dan $f_{\alpha/2}(v_1,v_2)$ adalah nilai dari distribusi F dengan v_1 dan v_2 degrees of freedom, yang berturut-turut menyisakan area sebesar 1-α/2 dan α/2 di sebelah kanan.

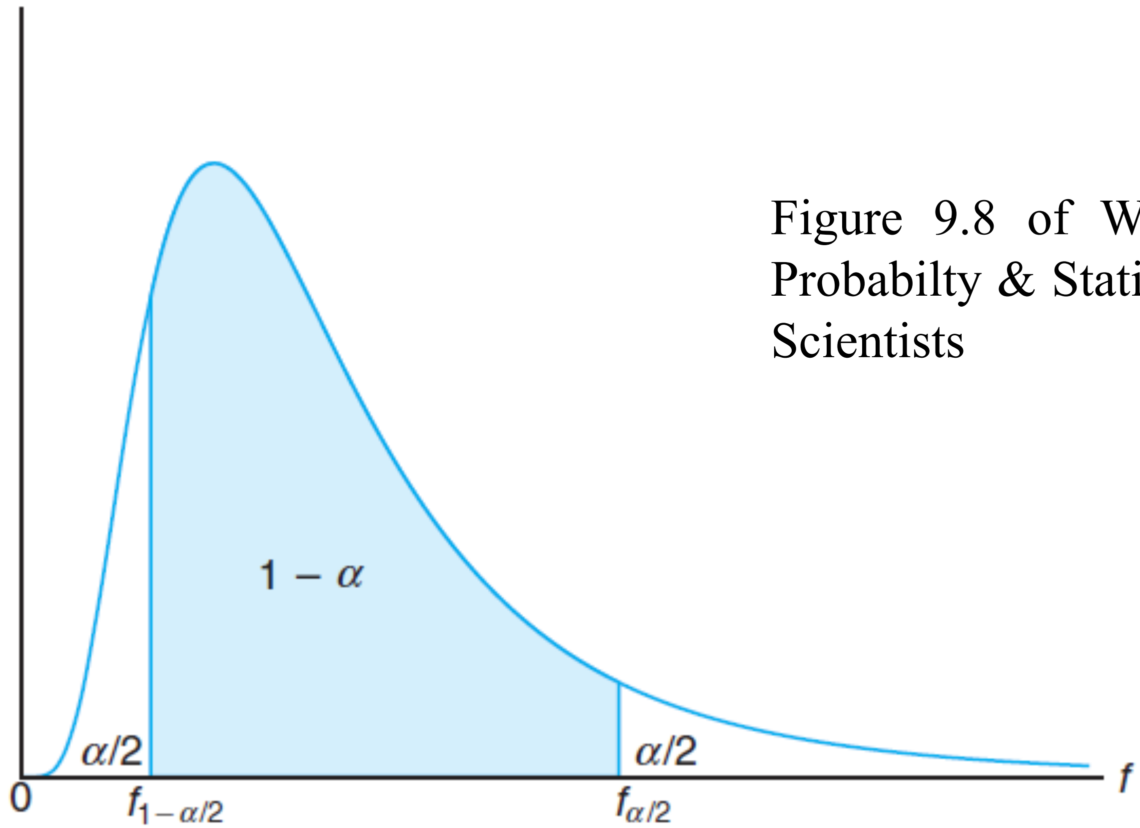


Figure 9.8 of Walpole, et al (2012), Probabilty & Statistics For Engineers & Scientists

Two Samples: Estimasi Rasio Dua Variances

- Sehingga diperoleh:

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

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

- Dengan menggunakan sifat distribusi F:

$$f_{1-\alpha/2}(v_1, v_2) = \frac{1}{f_{1-\alpha/2}(v_2, v_1)}$$

- Diperoleh:

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

Two Samples: Estimasi Rasio Dua Variances

Kesimpulan

- Jika s_1^2 dan s_2^2 adalah variance dari sampel-sampel yg independent berturut2 berukuran n_1 dan n_2 serta diambil dari populasi berdistribusi normal, maka 100(1-α)% confidence interval untuk σ_1^2/σ_2^2 adalah:

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

di mana $f_{\alpha/2}(v_1,v_2)$ adalah nilai f dengan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$ degrees of freedom yg menyisakan area sebesar $\alpha/2$ di sebelah kanan dan $f_{\alpha/2}(v_2,v_1)$ memiliki definisi serupa.

Contoh

- Perhatikan kembali tentang contoh uji senyawa fosfor di Sungai Brantas pada slide sebelumnya. Di contoh tersebut, confidence interval untuk selisih mean kandungan senyawa fosfor (mg/liter) pada 2 stasiun yang berbeda diperoleh dengan asumsi bahwa variance dari kedua populasi tidak sama. Dukung asumsi ini dengan mencari 98% confidence interval bagi $\sigma_1^2/\ \sigma_2^2$ dan σ_1/σ_2, jika σ_1^2 dan σ_2^2 berturut2 adalah variance populasi kandungan fosfor yang diperoleh pada stasiun 1 dan stasiun 2.

Contoh

Pada contoh uji senyawa fosfor di Sungai Brantas pada slide sebelumnya, diperoleh n_1=15, n_2=12, s_1=3.07, dan s_2=0.80. Untuk 98% confidence interval, α=0.02. Dengan menggunakan tabel distribusi F diperoleh $f_{0.01}(14,11) \approx 4.30$ dan $f_{0.01}(11,14) \approx 3.87$. Dgn demikian, 98% confidence interval untuk σ_1^2 / σ_2^2 dan σ_1/σ_2 adalah

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

$$3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$$

$$1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Tampak bahwa adalah keputusan tepat untuk berasumsi bahwa $\sigma_1^2 \neq \sigma_2^2$ dan $\sigma_1 \neq \sigma_2$ (1 tidak masuk di dalam interval)