



DISCRET RANDOM VARIABEL

Haniah Mahmudah

Discrete Random Variables

- Discrete Random Variables yg umum:
 - Bernoulli, Geometric, Binomial dan Poisson
 - **Bernoulli** – memodelkan eksperimen spt toss suatu coin
 - X adalah suatu indicator function
 - $X = 1 \rightarrow \text{sukses}; X = 0 \rightarrow \text{gagal}$
- $$f(x) = [P\{X=0\} = 1-p; P\{X=1\} = p]$$

$$\text{Mean} = p, \text{ Variance} = p(1-p)$$

Spt coin toss dg probabilitas p mendpkan *head*, $1-p$ mendpkan *tail*

Discrete Random Variables

- **Geometric** – memodelkan jumlah percobaan X sampai sukses pertama pd suatu deretan percobaan Bernoulli trials

$$P\{X = x\} = f(x) = (1-p)^{x-1}p;$$

dimana $x = 1, 2, 3, \dots$

$$\text{Mean} = 1/p$$

$$\text{Variance} = (1-p)/p^2$$

Sbg contoh, memodelkan jumlah *tail* yg terlihat sblm *head* pertama pd suatu deretan coin tosses

Discrete Random Variables

- **Binomial** – memodelkan jumlah sukses X pd n percobaan/trials. Mis p menyatakan probabilitas sukses pd 1 trial, probabilitas dari k sukses diberikan dg

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Mean = np , Variance = $np(1-p)$

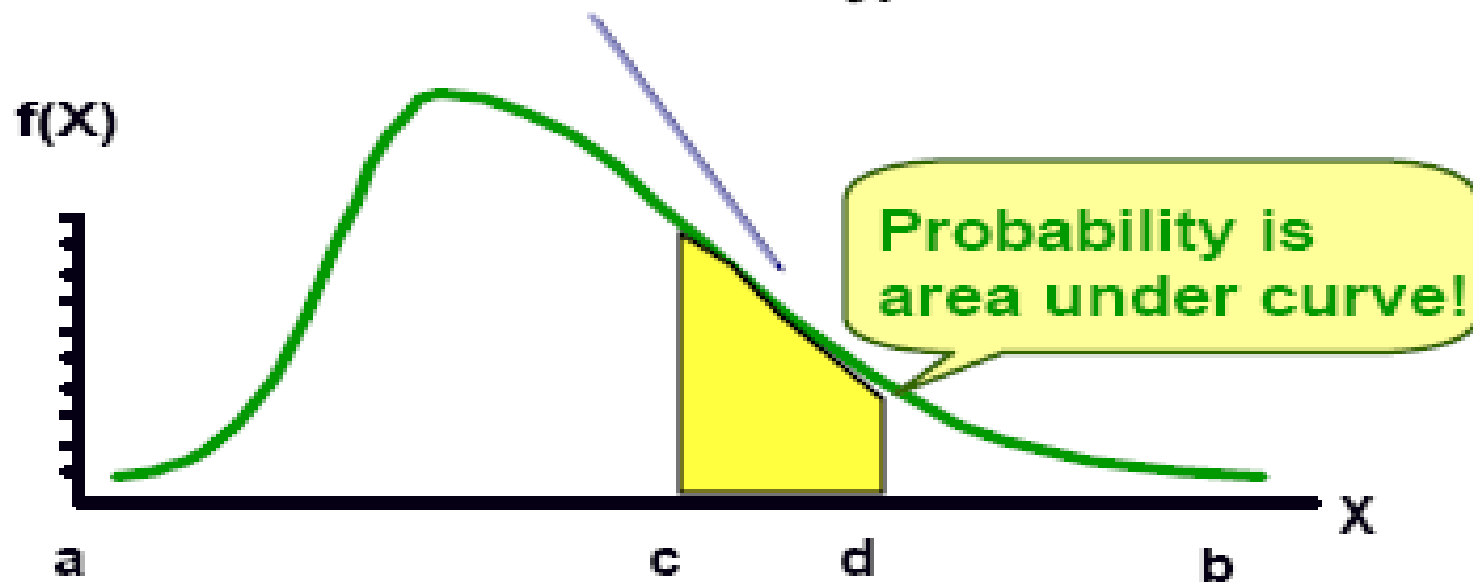
Tabel pd textbook memp macam-macam harga dari $P(X = k)$

Contoh Continuous Random Variable

Eksperimen	Random Variable	Harga Yg Mungkin
Berat mahasiswa ITB	Berat	43.2, 78, ... Kg
Umur hidup battery	Jam	900, 875.9, ... jam
Lama panggilan telepon	Lama panggilan	3.2, 1,53, ... menit
Waktu antar kedatangan paket ke router	Waktu antar kedatangan	0, 1.3, 2.78, ... det

Contoh Continuous Random Variable

$$P(c \leq X \leq d) = \int_c^d f(X) dx$$



Continuous Random Variable

- Continuous Random Variables yg umum:
 - Exponential, Uniform, Normal
- **Exponential** – memodelkan waktu antar kedatangan, lama waktu pelayanan (mis., waktu dari panggilan telepon), mis **X** suatu exponential

ra

$$f(x) = \frac{1}{a} e^{-x/a} \quad P\{X \leq x\} = F(x) = 1 - e^{-x/a}$$

$$\text{Mean} = a, \quad \text{Variance} = a^2$$

Continuous Random Variable

- **Uniform** – memodelkan kasus “equally likely”. Mis. X uniform random variable antara a dan b – yaitu X akan mempunyai harga antara a dan b dengan kemungkinan “equally likely”

$$f(x) = \frac{1}{b-a} \qquad F(x) = \frac{x-a}{b-a} \qquad a \leq x \leq b$$

$$\text{Mean} = (a+b)/2, \quad \text{Variance} = (b-a)^2/12$$

Continuous Random Variable

- **Normal** – Normal random variable memodelkan fenomena random alamiah utk jumlah yg besar. Mis **X** suatu normal random variable

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{- (x - \mu)^2 / 2\sigma^2}$$

$$\text{Mean} = \mu \qquad \text{Variance} = \sigma^2$$

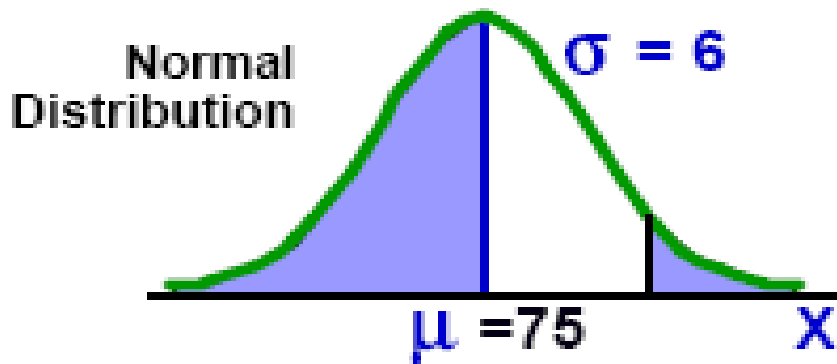
- **Standard Normal Z** adalah kasus dimana:
Mean = 0, Variance = 1.



Z Scores & Probability

- Normal Distribution
- Hubungan langsung antara persentase dan probabilitas
- Persentase dari kurva normal dp di- rephrased sbg problem probabilitas

Z Scores & Probability



- Berapakah probabilitas bhw pekerja pabrik yg dipilih random akan melaksanakan test dibawah 81 seconds atau diatas 75 seconds?

- Suatu konsultan menyelidiki waktu diperlukan pekerja pabrik utk assemble suatu part stlh mereka ditraining
- Konsultan menentukan bhw waktu dlm detik terdistribusi normal dg mean $\mu = 75$ seconds dan standard deviation $\sigma = 6$ seconds.

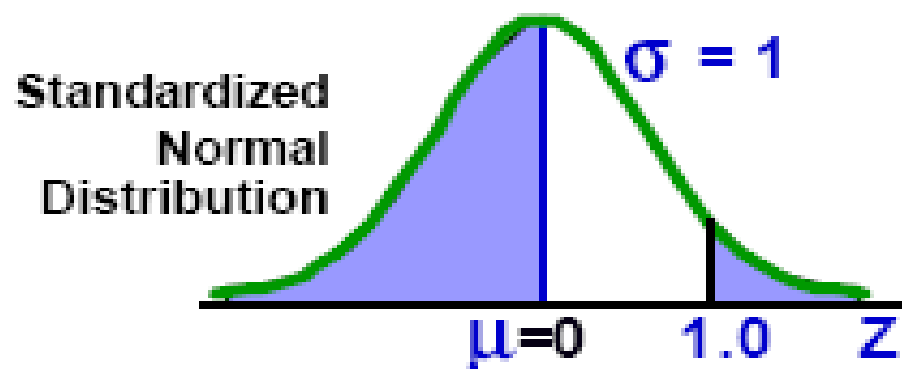
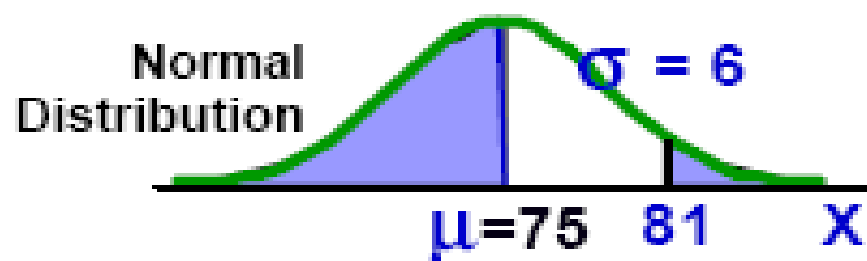
$$P(X < x) = P(Z < z)$$

$$\text{dimana } z = (x - \mu) / \sigma$$

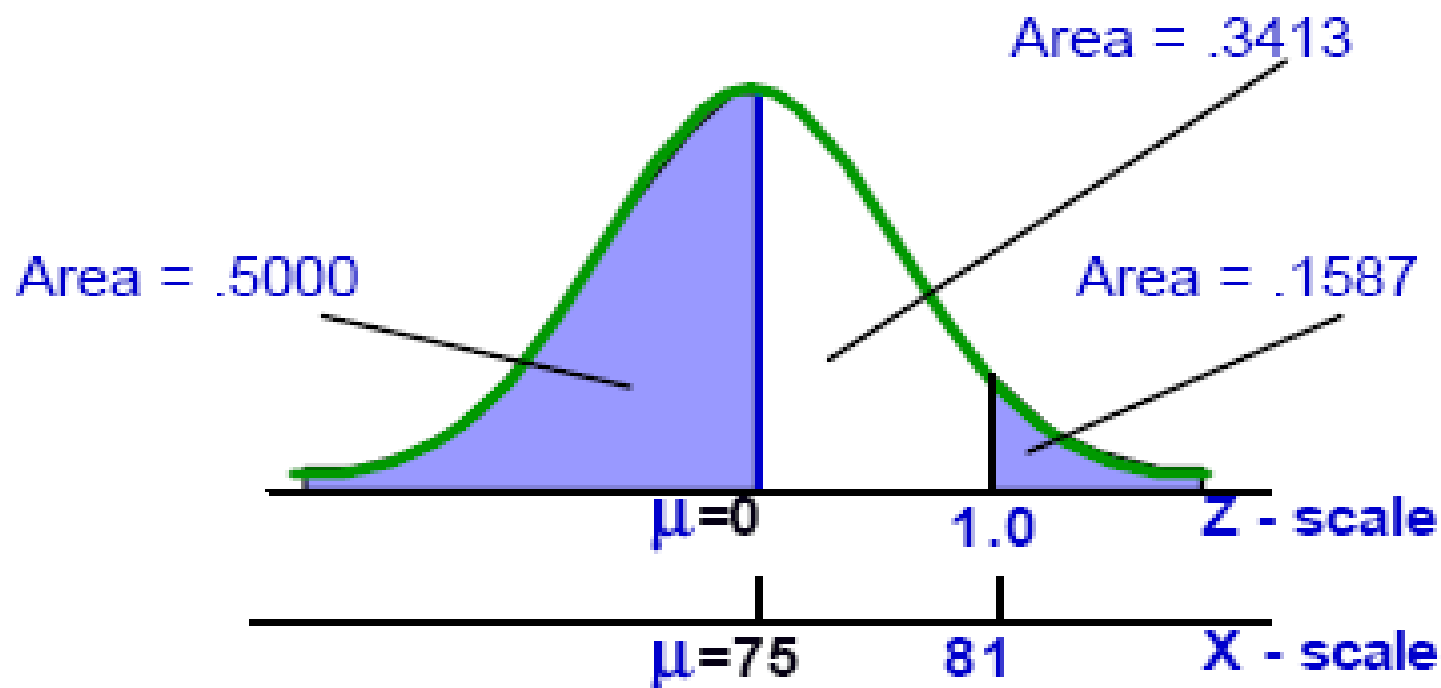
$$P(75 < X < 81)$$

$$Z = \frac{75 - 75}{6} = 0$$

$$Z = \frac{81 - 75}{6} = 1.0$$



$$P(75 < X < 81)$$



$$P(75 < X < 81) = .3413$$

Moments

- Ekspektasi $E[x]$ atau mean atau first moment dari suatu random variable X di definisikan dg

$$E[x] = \int_{-\infty}^{\infty} xf(x)dx \quad X \text{ continuous}$$

$$E[x] = \sum_{i \in S} x_i P\{X = x_i\} \quad X \text{ discrete}$$

Moment lebih tinggi didp dg mengganti x dg x^n

Variance, Mode, Quantile

- **Variance** didefinisikan sbg

$$\sigma^2 = E[x^2] - (E[x])^2$$

- **Mode** adalah titik dimana $f(x)$ adalah maximum
- **Quantile** – α quantile dari X ditulis x_α adalah titik pd X dimana $F(x_\alpha) = \alpha$
- Cat. 0,5 quantile disebut **median** dimana 50% harga pd kedua sisi

Aturan-Aturan untuk Random Variables

■ Aturan utk Means

- Suatu transformasi linier dari suatu random variable menghasilkan suatu linear scaling dari mean. Yaitu jika X adalah suatu random variable dg mean μ_X dan a dan b adalah konstanta maka jika $Y = aX + b$ mean dari Y diberikan oleh $\mu_Y = a\mu_X + b$
- Mean dari sum dari suatu set dari random variables adalah sum dari individual mean. Yaitu jika X dan Y adalah random variables maka $\mu_{X+Y} = \mu_X + \mu_Y$

Aturan-Aturan untuk Random Variables

■ Aturan utk Variances

- Suatu transformasi linier dari suatu random variable menghasilkan suatu squared scaling dari variance. Yaitu jika X adalah suatu random variable dg variance σ_x^2 dan a dan b adalah konstanta maka jika $Y = aX + b$ variance dari Y diberikan oleh $\sigma_y^2 = a^2 \sigma_x^2$
- Variance dari sum dari suatu set dari independent random variables adalah sum dari individual variances. Yaitu jika X dan Y adalah random variables maka $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$

Statistical Inference

- Menggunakan teori probabilitas utk membuat kesimpulan mengenai suatu populasi dari data sampel
- Tdk dp memperoleh data dari setiap anggota populasi maka menguji suatu **sampel random dari populasi** dan berdasarkan statistik dari sampel menyimpulkan mengenai parameter dari populasi



Statistical Inference

- Statistical Inference: menggunakan statistik dari suatu sampel random utk menyimpulkan mengenai parameter dari suatu populasi
 - Sbg contoh menguji mean \bar{x} dari sampel utk menyimpulkan mean dari populasi μ
 - Perlu mengerti bagaimana perubahan statistik dengan tiap sampel
- Sample Distribution: distribusi probabilitas dari suatu statistik (spt mean, standard deviation) dari semua sampel yg mungkin dari ukuran yg sama dari suatu populasi

Distribusi Sampel dari Counts dan Proportions

- Perhatikan suatu sampel random tetap (fixed) ukuran n dari observasi independen dari suatu populasi. Tiap observasi jatuh kedalam satu dari dua kategori, “sukses” atau “gagal”
 - Probabilitas suatu “sukses” (p) sama utk tiap observasi
 - Probabilitas suatu “gagal” ($1-p$)
- Mis X menyatakan *count* dari jumlah sukses dalam suatu sampel ukuran n . X memp distribusi Binomial

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Distribusi Sampel dari Counts dan Proportions

- Ingat distribusi Binomial memodelkan jumlah sukses X dlm n percobaan Bernoulli dan memp.

$$\text{Mean} = np, \text{ Variance} = np(1-p)$$

- Dg n bertambah besar distribusi dari X mendekati distribusi Normal dg mean dan variance

Distribusi Sampel dari Counts dan Proportions

- Utk estimasi probabilitas atau proportion dari suatu populasi p kita uji sample proportion:

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

dimana X adalah jumlah dari “sukses” dlm suatu sampel ukuran n

- \hat{p} adalah estimasi unbiased dari population proportion p .
- Jika ukuran sampel n besar, \hat{p} endekati suatu distribusi Normal dg

$$\text{Mean} = p$$

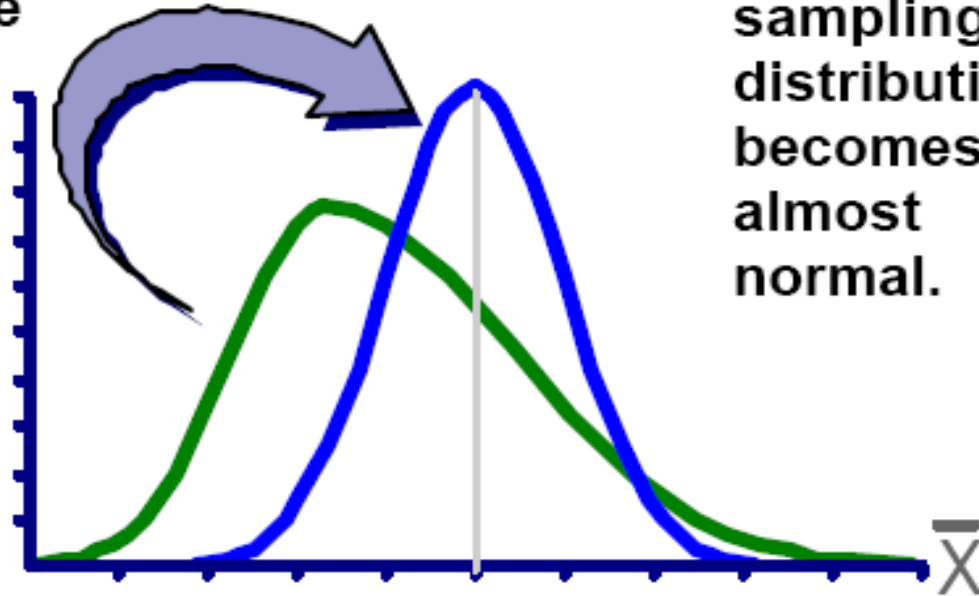
$$\text{Variance} = \frac{p(1-p)}{n}$$

Sample Distribution of Means

- Perhatikan suatu sampel random ukuran tetap n dari suatu populasi dg mean μ dan standard deviation σ . Distribusi dari sample mean \bar{x} (jika dihasilkan dari repeated random samples) memp. mean $= \mu$ dan standard deviation $= \sqrt{\frac{\sigma^2}{n}}$
- Jika populasi memp. distribusi Normal maka distribusi dari sample mean adalah Normal
- Dari Central Limit Theorem – distribusi dari suatu sum dari random variables mendekati distribusi Normal jika jumlah terms dlm sum menjadi besar

Central Limit Theorem

As sample
size gets
large
enough
(≥ 30) ...



sampling
distribution
becomes
almost
normal.

Central Limit Theorem

- Central limit theorem menyatakan bhw dg bertambah besarnya ukuran sampel n , tdk tergantung pd distribusi populasi, distribusi dari sample mean mendekati distribusi Normal utk ukuran sampel yg besa $\sqrt{\frac{\sigma}{n}}$ lg mean = μ dan standard deviation =

Tipe-Tipe Statistical Inference

- **Confidence Intervals**: mengestimasi harga suatu parameter populasi dg suatu harga rentang
 - Berapakah mean IQ dari mahasiswa SIT ITB?
 - Berapakah proporsi dari switches pd suatu network perlu perbaikan?
- **Hypothesis Testing**: menilai bukti yg disediakan data menyetujui suatu claim mengenai populasi
 - Apakah mean IQ dari mhs SIT ITB sama dg dg IQ populasi secara umum?
 - Apakah proporsi switches yg memerlukan perbaikan pd jaringan Telkom berbeda dg proporsi pd jaringan Indosat?

Point Estimation

- Menyediakan harga tunggal/single value, mis., sample mean, sample proportion
 - Berdasarkan observasi dari 1 sample
- Tdk memberikan informasi mengenai seberapa dekat harga point estimate thd parameter populasi yg tdk diketahui
- Contoh: Sample mean $\bar{X} = 22.9$ adalah point estimate dari mean populasi yg tdk diketahui μ

Interval Estimation

- Menyediakan nilai interval (a, b) dimana parameter populasi μ diprediksi berada
 - Interval berdasarkan observasi dari 1 sampel
- Memberikan informasi mengenai seberapa dekat dari estimasi ke parameter populasi yg tdk diketahui
 - Dp dinyatakan sbg
estimate \pm margin of error
 - Atau dinyatakan dlm terms probabilitas, (confidence level)

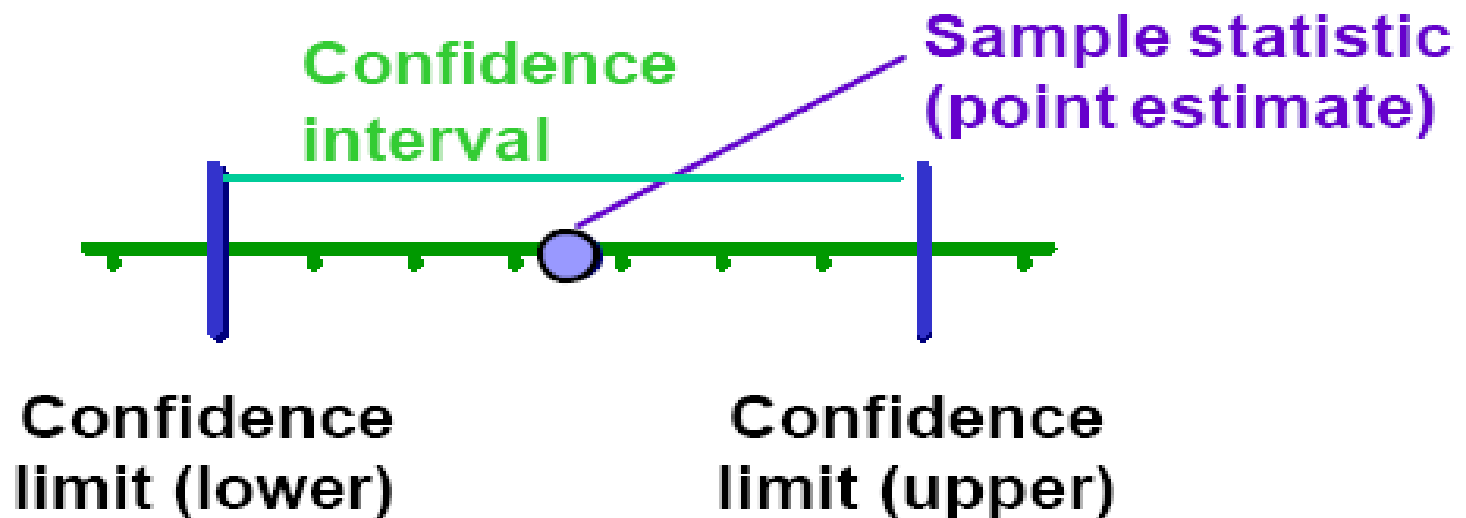
$$P(a \leq \mu \leq b) = 1 - \alpha$$

Level of Confidence

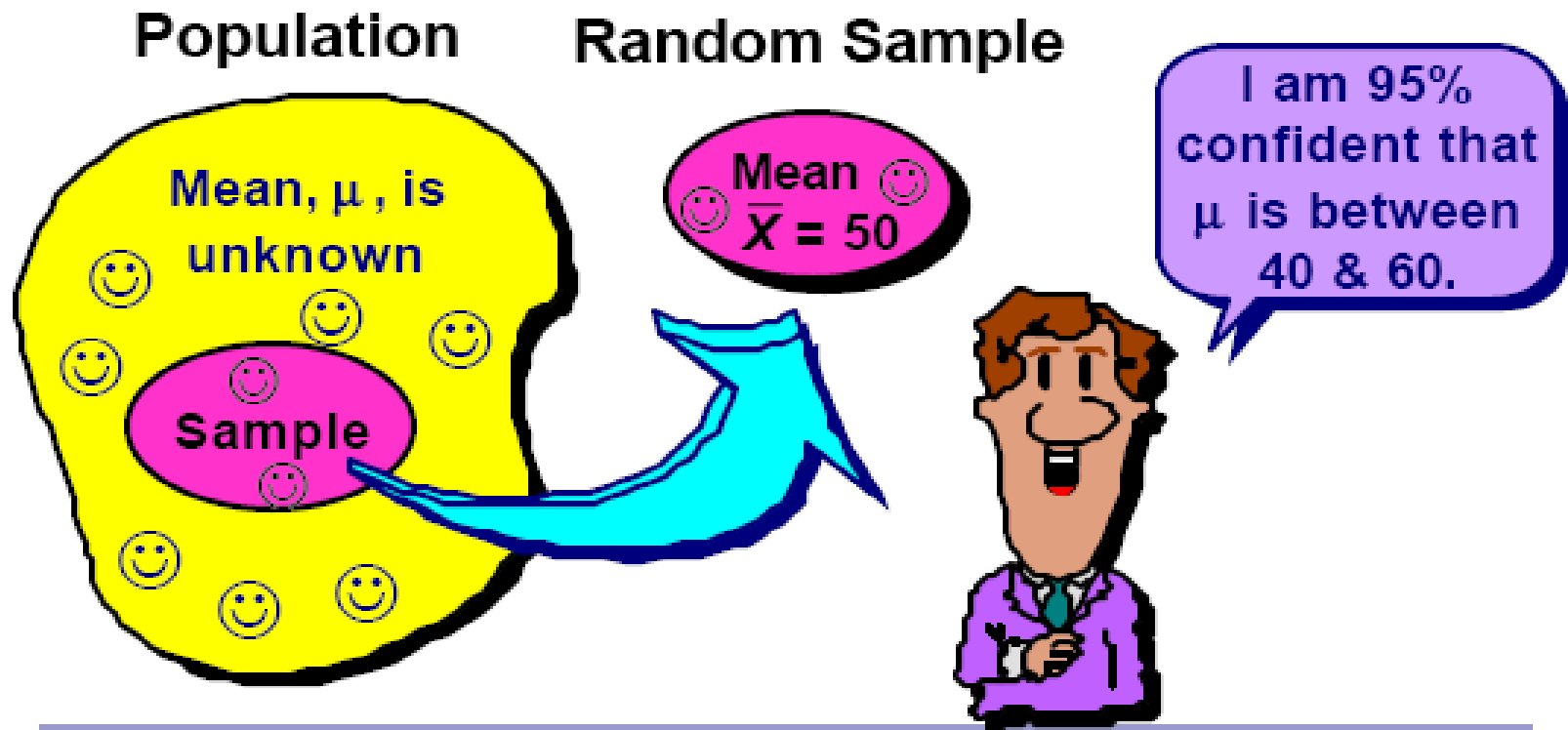
- Nilai α adalah probabilitas bhw parameter *tidak* berada dalam interval (a,b)
- $100(1 - \alpha) \%$ adalah confidence level dan adalah kemungkinan bhw parameter populasi yg tdk diketahui jatuh dlm interval (a,b)
- Nilai tipikal adalah $\alpha = .1, .05, .01$ yg memberikan confidence levels masing-masing 90%, 95%, dan 99%
- Contoh: Mean populasi yg tdk diketahui terletak antara 50 & 70 dg 95% confidence

Element Kunci dari Interval Estimation

A probability that the population parameter falls somewhere within the interval.



Confidence Interval Process



Confidence Interval utk Population Mean

- Asumsi

- Standard deviation populasi σ diketahui

- Ukuran sampel n cukup besar shg hasil central limit theorem dp diaplikasikan dan sample mean distribution dp diperkirakan dg distribusi normal. Aturan umum (Rule of thumb) utk ukuran sampel adalah ($n \geq 30$)

- $100(1-\alpha)$ % confidence interval pd sample mean diberikan oleh

$$\bar{x} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Confidence Interval for Population Mean

■ Catatan

- \bar{x} adalah sample mean.
- $Z_{(1-\alpha/2)}$ adalah nilai standard normal value dimana $\alpha/2$ adalah tail ke sebelah kanan dari nilai Z
- σ adalah standard deviation populasi
- n adalah ukuran sampel

$$\bar{x} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Contoh: Confidence Interval utk Population Mean

- Suatu retailer e-commerce spt Amazon.com, ingin melakukan studi waktu rata-rata (mean time) yg diperlukan utk memproses dan mengapalkan pesanan. Suatu random sample dari waktu utk proses dan mengapalkan 33 pesanan dikumpulkan dan dinyatakan sbg n dlm jam di bawah. Dari data yg lalu standard deviation dari populasi $\sigma = 9$

{23, 33, 14, 15, 42, 28, 33, 45, 23, 34, 39, 21, 36, 23, 34, 36, 25, 9, 11, 19, 35, 24, 31, 29, 16, 23, 34, 24, 38, 15, 13, 35, 28}

- Tentukan 90% confidence interval dari rata-rata waktu proses dan pengapalan pesanan.
 - sample mean \bar{x} adalah = 26.9091, ukuran sampel $n = 33$ pd 90% confidence level $Z_{(1-\alpha/2)} = Z_{.95} = 1.645$

Contoh: Confidence Interval utk Population Mean

- Krnnya confidence interval adalah menghasilkan

$$\bar{x} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$26.9091 \pm 1.645 \left(\frac{9}{\sqrt{33}} \right)$$

$$26.9091 \pm 2.577$$

Cat margin of error kadang-kadang diekspresikan sbg persentase dari estimasi. Utk contoh e-commerce:

$$\text{margin of error \%} = 100 * (2.577 / 26.9091) = 9.57\%$$

- Juga confidence interval dp diekspresikan sbg (24.332, 29.486) yg dp diinterpretasikan sbg

$$P(24.332 < \mu < 29.486) = .9$$

Confidence Intervals

- Trade off antara confidence level $100(1-\alpha) \%$ dan margin of error
 - Lebih tinggi confidence \rightarrow lebih tinggi harga $Z \rightarrow$ lebih besar margin of error
- Contoh proses dan pengiriman pemesanan e-commerce. Suatu 95% confidence interval memp. $Z = 1.96$ (dimana $26.9091 \pm 1.96(\frac{9}{\sqrt{33}}) 1.645$) dan sbg hasil
$$26.9091 \pm 3.0707$$

Confidence Intervals

- Margin of error juga tergantung pd ukuran sampel n , lebih besar n makin kecil margin of error
- Utk confidence interval pd population mean, margin of error berkurang setengahnya utk tiap pertambahan faktor 4 pd ukuran sampel

$$\bar{x} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Utk contoh e-commerce interval ukuran sample adalah 4 kali lebih besar (yaitu 132) dg mean dan standard deviation yg sama interval akan

$$26.9091 \pm 1.2885$$

Confidence Intervals

- Cat utk margin of error yg diinginkan m kita dp tentukan ukuran sampel yg diperlukan n utk mencapai m . Kita mendpkan

$$n = \left(\frac{Z \sigma}{m \cdot \bar{x}} \right)^2$$

- Utk contoh e-commerce pd 90% confidence level jika diinginkan margin of error 3%, $m \cdot \bar{x} = .03 \times 26.9091 = .80727$ dan selesaikan utk ukuran sampel n

$$n = \left(\frac{Z\sigma}{m \cdot \bar{x}} \right)^2 = \left(\frac{1.645 \times 9}{.80727} \right)^2 = 336.34 \Rightarrow 337$$

Cat perlu $337 - 33 = 304$ tambahan observasi

Confidence Interval utk Proportion of population

- Dari aproksimasi Normal pd distribusi Binomial kita dapatkan $100(1 - \alpha)\%$ confidence interval pd suatu population proportion sbg

$$\hat{p} \pm Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

dimana $Z_{1-\alpha/2}$ adalah $\alpha/2$ critical point dari standard distribusi Normal

Confidence Interval utk Proportion of population

- Contoh: Perhatikan suatu link komunikasi satelit. Spy dp mengestimasi packet error rate pd link kita transmit 5000 packets dan observasi bhw 23 diterima error. Tentukan 90% confidence interval pd packet error probability. Dari, $Z_{.95} = 1.645$, $n = 5000$,

$$\hat{p} = 23/5000 = .0046$$

$$\hat{p} \pm Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .0046 \pm .0016$$

- Krnnya 90% confidence interval utk packet error probability diberikan oleh (.0030, .0062)

Confidence Interval utk Quantile of population

- Quantile: Harga x_q dimana CDF mempunyai harga q disebut q -quantile atau $100-q$ -percentile

$$P(x \leq x_q) = F(x_q) = q \quad ; 0 \leq q \leq 1$$

- 50-percentile (atau 0.5-quantile) disebut median
- Posisi dari suatu harga q -quantile value dari suatu sorted order list $x_1, x_2, x_3, \dots, x_n$ adalah

$$\left[(n - 1)q + 1 \right]$$

* dibulatkan ke integer terdekat

Confidence Interval utk Quantile of population

- 100(1- α)% confidence interval pd suatu harga populasi q -quantile x_q adalah

$$P(y_l \leq x_q \leq y_u) \approx 1 - \alpha$$

dima
$$l = \lfloor nq + 0.5 - Z_{1-\alpha/2} \sqrt{nq(1-q)} \rfloor$$

$$u = \lceil nq + 0.5 + Z_{1-\alpha/2} \sqrt{nq(1-q)} \rceil$$

Confidence Interval utk Quantile of population

- Contoh: 45 titik data ($n=45$)
- 6, 6, 7, 8, 8.5, 9, 11, 13, 15, 24, 29, 30, 32, 34, 37, 39, 41, 42, 42, 43, 46, 47, 47.5, 49, 50, 52, 54, 55, 59, 62, 63, 66, 68, 71, 81, 83, 84, 88, 93, 97, 103, 108, 111, 116, 134

Cari 90% c.i. pd 0.5 quantile.

Posisi dari 0.5 quantile = $(45-1)*0.5 + 1 = 23 \rightarrow x_{0.5} = 47.5$

$$l = \lfloor 17.4825 \rfloor = 17$$

$$u = \lceil 28.5175 \rceil = 29$$

- Krnnya, 90% c.i. pd $x_{0.5} = (41, 59)$

Tugas (kumpulkan minggu depan)

1. Perhatikan N mobile phones dlm suatu cell. Tiap phone mungkin berusaha utk transmit data pd suatu kanal shared time slot. Tiap transmisi terjadi tepat pd satu slot, dan tdk ada pencegahan collision digunakan serta tiap phone akan transmit dlm suatu slot dg probabilitas p , independen thd phone lainnya.
 - a). Berapakah probabilitas suatu time slot kosong, yaitu tdk ada usaha dari sembarang phone?
 - b). Berapakah probabilitas suatu transmisi sukses, yaitu secara tepat satu phone berusaha transmit.
 - c). Berapakah probabilitas collision pd suatu slot, yaitu dua atau lebih phone berusaha transmit pd slot yg sama?

Tugas (cont.)

2. Dlm suatu access switched data network, user bisa request suatu koneksi utk di-set up utk suatu transfer data. Jika suatu call-setup request tiba, suatu access network node akan menentukan apakah menerima permintaan atau menolak berdasarkan ketersediaan resources. Jika permintaan ditolak, user akan mengulang usaha sampai 10 kali sblm menyerah. Asumsikan bhw tiap permintaan call-setup memp. probabilitas 0.02 utk diterima dan usaha permintaan panggilan adalah independent.
- a). Berapakah probabilitas suatu permintaan panggilan diberikan pd usaha pertama?
 - b). Berapakah probabilitas bhw suatu panggilan pertama diterima adalah usaha yg ke-empat?
 - c). Berapakah probabilitas bhw memerlukan lebih dari lima usaha bagi user utk koneksi?
 - d). Berapakah probabilitas bhw user akhirnya menyerah?
 - e). Berapakah rata-rata jumlah usaha panggilan diperlukan utk koneksi?

Tugas (cont.)

3. Perhatikan waktu transaksi pd suatu aplikasi web-based, dari 3000 transaksi terdistribusi normal dg mean 66 sec dan standard deviation 3 sec. jika 80 sampel masing-masing terdiri dari 25 transaksi didapat,
- a). berapakah mean dan standard deviation yg diharapkan sbg hasil dari mean dari distribusi sampling?
 - b). Dalam berapa banyak sampel kita bisa mengharapkan mean: (i) antara 64.8 dan 66.3 sec dan (ii) kurang dari 64.4 sec?