

Kursus Statistika Dasar

Bambang Suryoatmono

Bagian 1

Statistika Deskriptif

Pengelompokan Statistika

- **Statistika Deskriptif:** statistika yang menggunakan data pada suatu kelompok untuk menjelaskan atau menarik kesimpulan mengenai kelompok itu saja
 - Ukuran Lokasi: *mode, mean, median*, dll
 - Ukuran Variabilitas: *varians, deviasi standar, range*, dll
 - Ukuran Bentuk: *skewness, kurtosis, plot boks*
- **Statistika Inferensi (Statistika Induksi):** statistika yang menggunakan data dari suatu sampel untuk menarik kesimpulan mengenai populasi dari mana sampel tersebut diambil

Pengelompokan Statistika lainnya

- **Statistika Parametrik:**
 - Menggunakan asumsi mengenai populasi
 - Membutuhkan pengukuran kuantitatif dengan level data interval atau rasio
- **Statistika Nonparametrik (*distribution-free statistics for use with nominal / ordinal data*):**
 - Menggunakan lebih sedikit asumsi mengenai populasi (atau bahkan tidak ada sama sekali)
 - Membutuhkan data dengan level serendah-rendahnya ordinal (ada beberapa metode untuk nominal)

Istilah-istilah Dasar

- **Populasi:** sekumpulan orang atau objek yang sedang diteliti
- **Sensus:** pengumpulan data pada *seluruh* populasi
- **Sampel:** sebagian dari populasi yang, apabila diambil dengan benar, merupakan representasi dari populasi
- **Parameter:** ukuran deskriptif dari populasi
- **Statistik:** ukuran deskriptif dari sampel

Jenis Data

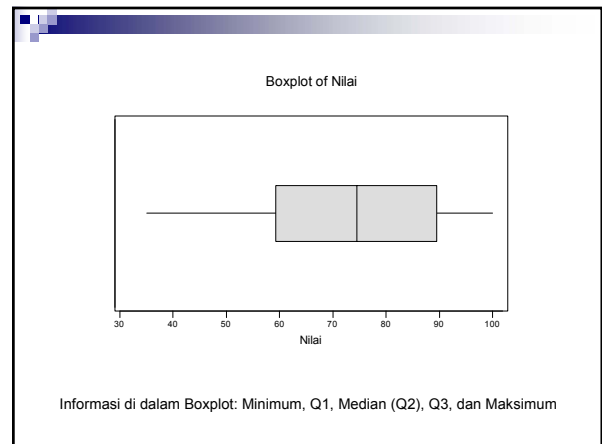
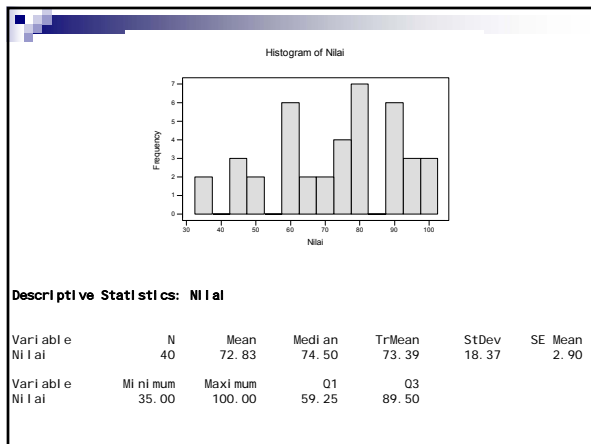
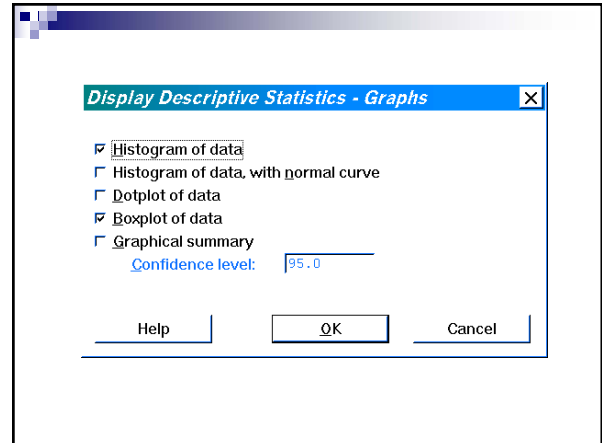
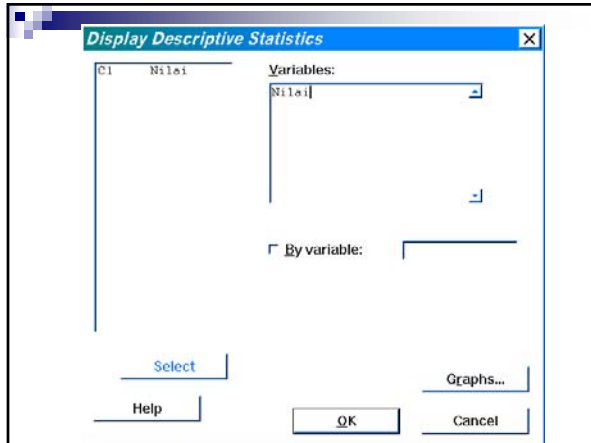
	Nominal	Ordinal	Interval	Rasio
Bilangan menunjukkan perbedaan	√	√	√	√
Pengukuran dapat digunakan untuk membuat peringkat atau mengurutkan objek		√	√	√
Perbedaan bilangan mempunyai arti			√	√
Mempunyai nol mutlak dan rasio antara dua bilangan mempunyai arti				√

Distribusi Frekuensi

- *Ungrouped Data vs Grouped data*
- *Range*
- *Class midpoint*
- *Frekuensi Relatif*
- *Frekuensi Kumulatif*

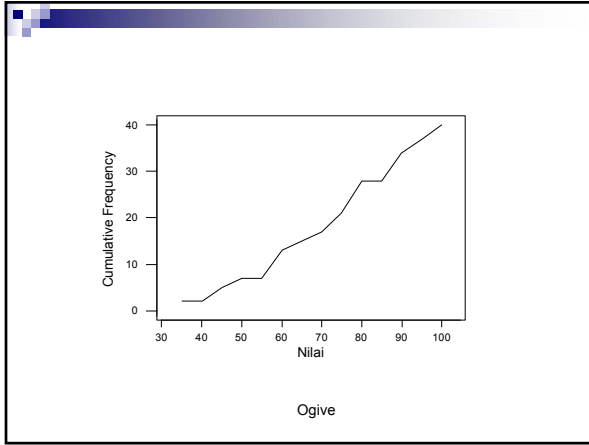
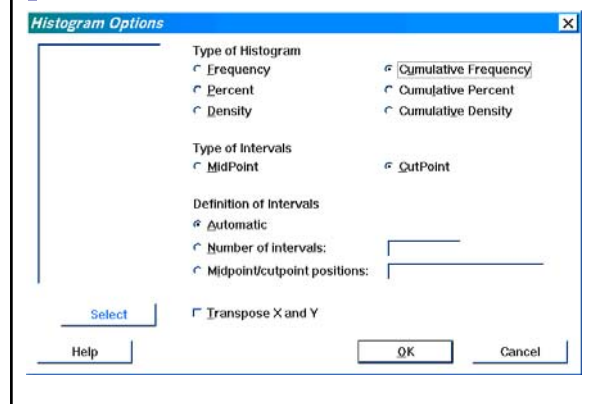
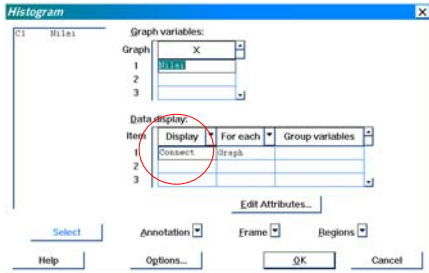
Histogram (contoh MINITAB)

- **Data:**
60, 65, 70, 73, , (tulisi di C1)
- **MINITAB:**
Stat -> Basic Statistics -> Display Descriptive Statistics



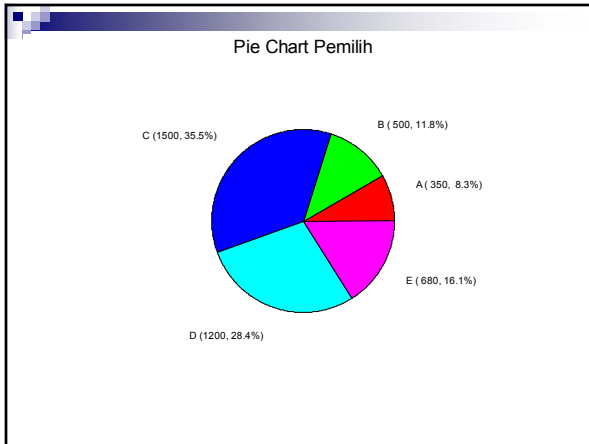
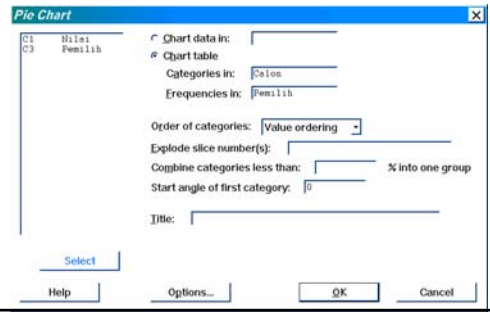
Ogive (Poligon Frekuensi Kumulatif)

MINITAB:
Graph -> Histogram



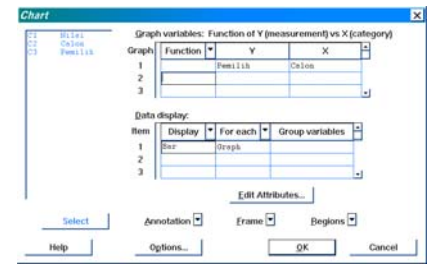
Pie Chart

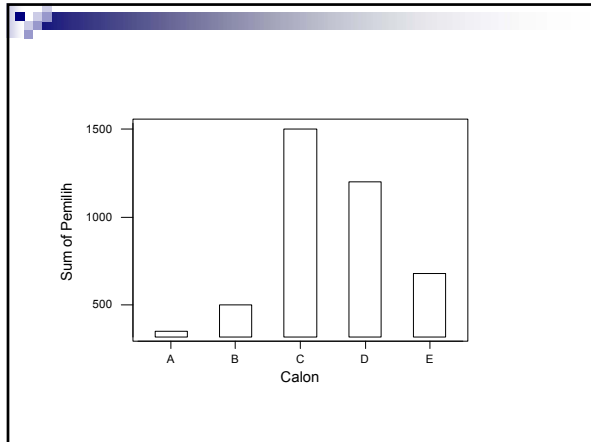
MINITAB: Graph -> Pie Chart



Bar Chart

MINITAB: Graph ->
Chart





Stem-and-leaf

- MINITAB: Graph -> Character Graph -> Stem-and-leaf

Character Stem-and-Leaf Display

Stem-and-Leaf of HrgTanah N = 50
Leaf Unit = 10

```

1  4 5
2  5 5
3  6 5
7  7 2699
20 8 001245555669
(18) 9 00000034555556779
12 10 0000
8  11 0000
4  12 0
3  13 00
1  14
1  15 0

```

Ukuran Lokasi pada data tak terkelompok

- Mean = rata-rata hitung = rata-rata
 μ = rata-rata populasi, \bar{x} = rata-rata sampel
- Median = nilai tengah dari data yang diurutkan
- Mode = nilai yang paling sering terjadi pada suatu data
- Persentil = ukuran lokasi yang membagi sekelompok data menjadi 100 bagian
- Quartil = ukuran lokasi yang membagi sekelompok data menjadi 4 bagian atau subkelompok

Ukuran Lokasi pada data tak terkelompok (lanjutan)

Mencari persentil ke p:

- Urutkan n data dari kecil ke besar
- Hitung lokasi persentil $i = (p/100) * n$
- Jika $i =$ bil bulat, maka persentil ke p adalah (bil ke $i +$ bil ke $i+1$) / 2
- Jika i bukan bil bulat, maka persentil ke p adalah bil ke $\text{int}(i) + 1$

Ukuran Variabilitas pada data tak terkelompok

- Range = maksimum – minimum
- Interquartile range = $Q_3 - Q_1$
 $Q_3 =$ persentil ke 75, $Q_1 =$ persentil ke 25
- Deviasi absolut rata-rata $MAD = \frac{\sum |X - \mu|}{N}$
- Varians populasi

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

Ukuran Variabilitas pada data tak terkelompok (lanjutan)

- Varians sampel:

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

- Catatan: N = ukuran populasi, n = ukuran sampel

Ukuran Variabilitas pada data tak terkelompok (lanjutan)

- Deviasi Standar Populasi

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

- Deviasi Standar Sampel

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

- Note: Koefisien Variasi $CV = \frac{\sigma}{\mu} 100\%$

Ukuran Lokasi pada data terkelompok

- Rata-rata

$$\mu_{grouped} = \frac{\sum fM}{\sum f} = \frac{\sum fM}{N}$$

f = frekuensi kelas
N = frekuensi total

Ukuran Variabilitas pada data terkelompok

- Varians populasi dan deviasi standar populasi

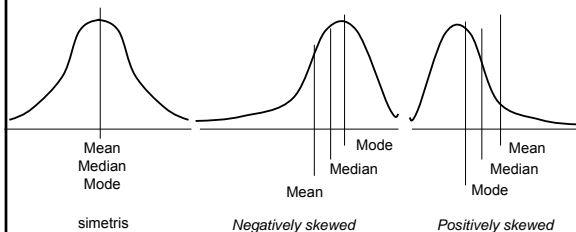
$$\sigma^2 = \frac{\sum f(M - \mu)^2}{N} = \frac{\sum fM^2 - \frac{(\sum fM)^2}{N}}{N}; \sigma = \sqrt{\sigma^2}$$

- Varians sampel dan deviasi standar sampel

$$S^2 = \frac{\sum f(M - \bar{X})^2}{n-1} = \frac{\sum fM^2 - \frac{(\sum fM)^2}{n}}{n-1}; S = \sqrt{S^2}$$

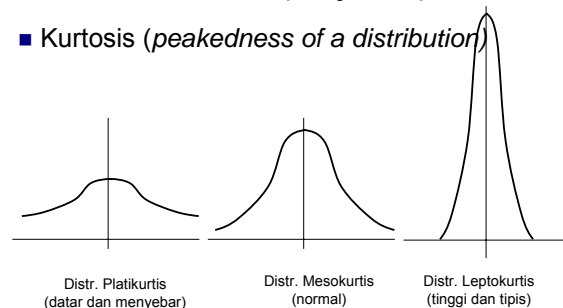
Ukuran Bentuk

- Skewness



Ukuran Bentuk (lanjutan)

- Kurtosis (*peakedness of a distribution*)



Bagian 2

Probabilitas

Probabilitas

- $P(A)$ = peluang (probabilitas) bahwa kejadian A terjadi
- $0 \leq P(A) \leq 1$
- $P(A) = 0$ artinya A pasti terjadi
- $P(A) = 1$ artinya A tidak mungkin terjadi
- Penentuan nilai probabilitas:
 - Metode Klasikal
 - Menggunakan Frekuensi Relatif Kejadian
 - Dengan carai subyektif

Metode Klasikal untuk menentukan Probabilitas

- Metode ini menggunakan:
 - Eksperimen, yaitu proses yang menghasilkan *outcome*, dan
 - *Event*, yaitu *outcome* dari suatu eksperimen

$$P(E) = \frac{n_e}{N}$$

- N = total banyaknya *outcome* yang mungkin pada suatu eksperimen
- n_e = banyaknya *outcome* di mana *event* E terjadi
- Pada metode ini probabilitas dapat ditentukan sebelum eksperimen dilakukan (*a priori*)

Metode Frekuensi Relatif Kejadian untuk Menentukan Probabilitas

- Pada metode ini probabilitas suatu event didapat dari banyaknya *event* tersebut terjadi di masa lalu, dibagi dengan banyak total kesempatan *event* tersebut terjadi.

Probabilitas Subyektif

- Hanya didasarkan atas perasaan, intuisi, atau pengetahuan orang yang menentukan probabilitas
- Meskipun bukan merupakan cara yang ilmiah, namun pendekatan ini dapat saja menghasilkan probabilitas yang cukup akurat

Struktur Probabilitas

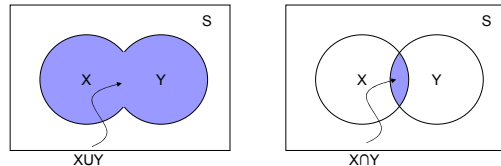
- Eksperimen. Contoh: Mencatat kurs US\$ terhadap rupiah setiap hari Senin pukul 9 pagi selama 12 bulan
- *Event*. Contoh: mendapati kurs US\$ terhadap rupiah kurang dari 10000
- *Elementary Event*: adalah *event* yang tidak dapat dipecah lagi menjadi *event* lain.
- Ruang sampel (*sample space*): adalah daftar atau tabel lengkap yang memuat semua *elementary event* pada suatu eksperimen.

Struktur Probabilitas (lanjutan)

- Contoh Ruang Sampel:
Wawancara dengan pertanyaan jenis penanaman modal (PMA atau PMDN), maka ruang sampelnya adalah:
 - Utk 1 responden: {PMA, PMDN}
 - Utk 2 responden: {PMA-PMA, PMA-PMDN, PMDN-PMA, PMDN-PMDN}

Struktur Probabilitas (lanjutan)

- *Union* = "atau" = gabungan. Simbol: \cup .
- *Intersection* = "dan" = irisan. Simbol: \cap .
- Contoh: Jika diketahui $X = \{1, 4, 7, 9\}$ dan $Y = \{2, 3, 4, 5, 6\}$, maka
 - $X \cup Y = \{1, 2, 3, 4, 5, 6, 7, 9\}$
 - $X \cap Y = \{4\}$



Struktur Probabilitas (lanjutan)

- *Mutually Exclusive Events*: adalah kejadian-kejadian yang tidak mempunyai irisan. Artinya, kejadian yang satu meniadakan kejadian yang lainnya; kedua kejadian tidak dapat terjadi secara simultan. Jadi:

$$P(X \cap Y) = 0$$

apabila X dan Y *mutually exclusive*.

Struktur Probabilitas (lanjutan)

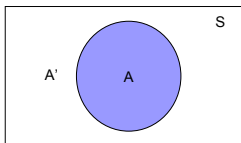
- *Independent Events*: adalah kejadian-kejadian satu sama lain tidak saling mempengaruhi. Artinya, terjadi atau tidak terjadinya satu kejadian tidak mempengaruhi terjadinya atau tidak terjadinya kejadian yang lainnya. Jadi:

$$P(X|Y) = P(X) \text{ dan } P(Y|X) = P(Y)$$

apabila X dan Y *adalah* kejadian independen. $P(X|Y)$ artinya probabilitas bahwa X terjadi apabila diketahui Y telah terjadi.

Struktur Probabilitas (lanjutan)

- *Collectively Exhaustive Events*: adalah daftar semua kejadian elementer (*elementary events*) yang mungkin terjadi pada sebuah eksperimen. Jadi sebuah ruang sampel selalu terdiri atas *Collectively Exhaustive Events*.
- Komplemen dari kejadian A, diberi notasi A' yang artinya "bukan A" adalah semua kejadian elementer pada suatu eksperimen yang bukan A. Jadi: $P(A) + P(A') = 1$



Aturan hitungan mn

- Untuk suatu operasi yang dapat dilakukan dengan m cara dan operasi ke dua yang dapat dilakukan dengan n cara, maka kedua operasi dapat terjadi dalam mn cara. Aturan ini dapat dikembangkan untuk tiga atau lebih operasi.

Pengambilan Sampel dari Suatu Populasi

- Pengambilan sampel berukuran n dari dari populasi berukuran N dengan penggantian (*with replacement*) akan menghasilkan N^n kemungkinan
- Pengambilan sampel berukuran n dari dari populasi berukuran N tanpa penggantian (*without replacement*) akan menghasilkan

$${}_N C_n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

kemungkinan

Marginal, Union, Joint, and Conditional Probabilities

- *Marginal Probability*: $P(A)$ = probabilitas bahwa A terjadi
- *Union Probability*: $P(A \cup B)$ = probabilitas bahwa A atau B terjadi
- *Joint Probability*: $P(AB) = P(A \cap B)$ = probabilitas bahwa A dan B terjadi
- *Conditional Probability*: $P(A|B)$ = probabilitas bahwa A terjadi apabila diketahui B telah terjadi

Aturan Perjumlahan

- Aturan Umum Perjumlahan:

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

- Aturan Khusus Perjumlahan:
Apabila X dan Y adalah kejadian yang *mutually exclusive*, maka

$$P(X \cup Y) = P(X) + P(Y)$$

Aturan Perkalian

- Aturan Umum Perkalian:

$$P(X \cap Y) = P(X) * P(Y|X) = P(Y) * P(X|Y)$$

- Aturan Khusus Perkalian:
Apabila X dan Y adalah kejadian yang independen, maka

$$P(X \cup Y) = P(X) * P(Y)$$

Aturan untuk Probabilitas Bersyarat (*Conditional Probability*)

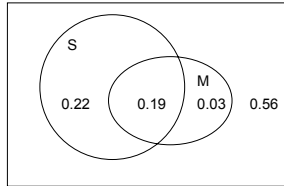
- Probabilitas bahwa X terjadi apabila diketahui Y telah terjadi

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(X) * P(Y | X)}{P(Y)}$$

Contoh Soal tentang Probabilitas

- Di sebuah kota, diketahui bahwa:
 - 41% penduduk mempunyai sepeda motor
 - 19% mempunyai sepeda motor dan mempunyai mobil
 - 22% mempunyai mobil
- Apakah kepemilikan sepeda motor dan kepemilikan mobil di kota tersebut independen? Gunakan data di atas untuk menjawabnya.
- Bila seorang penduduk di kota tersebut diambil secara acak berapa probabilitas bahwa ia memiliki sepeda motor dan tidak memiliki mobil?
- Bila seorang penduduk di kota tersebut diambil secara acak dan diketahui ia memiliki mobil, berapa probabilitas bahwa ia tidak memiliki sepeda motor?
- Bila seorang penduduk di kota tersebut diambil secara acak, berapakah probabilitas bahwa ia tidak memiliki sepeda motor dan tidak memiliki mobil?

Jawab



- S = memiliki sepeda motor; M = memiliki mobil
 $P(S) = 0.41$, $P(SM) = 0.19$, $P(M) = 0.22$.
 - Karena $P(S)P(M) \neq P(SM)$, maka kepemilikan sepeda motor dan kepemilikan mobil tidak independen.
 - dengan diagram Venn didapatkan $P(SM') = 0.22$.
 - $P(S'|M) = P(S'M) / P(M) = 0.03 / 0.22 = 0.1364$
 - $P(S'M') = 0.56$ (dari diagram Venn)

Contoh Soal tentang Probabilitas

- Hasil sebuah survei yang menanyakan "Apakah Anda mempunyai komputer dan/atau kalkulator di rumah?" adalah sebagai berikut. Apakah kepemilikan kalkulator dan kepemilikan komputer independen?

		Kalkulator	
		Ya	Tdk
Komputer	Ya	46	3
	Tdk	11	15

Jawab

A = memiliki komputer; B = memiliki kalkulator

		Kalkulator		
		Ya	Tdk	
Komputer	Ya	46	3	49
	Tdk	11	15	26
		57	18	75

$$P(A) = \frac{49}{75} = 0.653$$

$$P(B) = \frac{57}{75} = 0.76$$

$$P(AB) = \frac{46}{75} = 0.613$$

$$P(A) * P(B) = 0.653 * 0.76 = 0.49628 \neq P(AB)$$

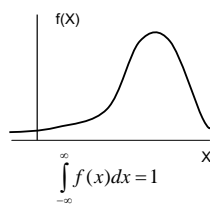
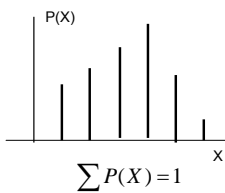
→ A dan B tidak independen

Bagian 3

Variabel Acak Diskret

Variabel Acak (X)

- Variabel acak diskret: X hanya mempunyai sejumlah terbatas nilai
- Variabel acak kontinu: X dapat mempunyai tak hingga nilai



Variabel Acak Diskret

- Rata-rata

$$\mu = E(X) = \sum X * P(X)$$

- Deviasi Standar

$$\sigma = \sqrt{\sum (X - \mu)^2 * P(X)}$$

Distribusi Binomial

- Distribusi Binomial** $P(X) = \frac{n!}{X!(n-X)!} p^X q^{n-X}$

n = # trials
 X = # sukses
 p = probabilitas sukses pada satu trial
 $q = 1 - p$ = probabilitas gagal pada satu trial

- Rata-rata Distribusi Binomial** $\mu = n * p$
- Deviasi Standar Distribusi Binomial** $\sigma = \sqrt{n * p * q}$
- Terjadi pada: eksperimen yang terdiri atas n trials, dengan setiap trial mempunyai probabilitas sukses p (konstan)
- MINITAB: Calc -> Probability Distribution -> Binomial

Distribusi Binomial (lanjutan)

Distribusi Poisson

- Distribusi Poisson:**

$$P(X) = \frac{\lambda^X e^{-\lambda}}{X!}$$

$X = 0, 1, 2, \dots$
 λ = rata-rata
 $e = 2.718282$

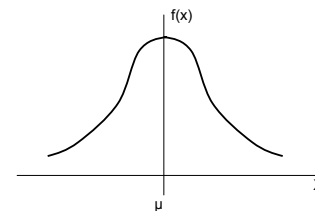
- Rata-rata Distribusi Poisson** $\mu = \lambda$
- Deviasi standar Distribusi Poisson** $\sigma = \sqrt{\lambda}$
- Distribusi Poisson merepresentasikan kejadian yang amat jarang. X = banyaknya kejadian tersebut terjadi pada suatu waktu atau area
- MINITAB: Calc -> Probability Distribution -> Poisson

Bagian 4

Variabel Acak Kontinu

Distribusi Normal dan Normal Standar

- Distribusi Normal (=Gauss)**
 Parameter: μ = rata-rata, dan σ = deviasi standar



Distribusi Normal dan Normal Standar (lanjutan)

- Distribusi Normal Standar = distribusi normal untuk $\mu = 0$ dan $\sigma = 1$. Konversi dari X yang terdistribusi normal ke Z yang terdistribusi normal standar:

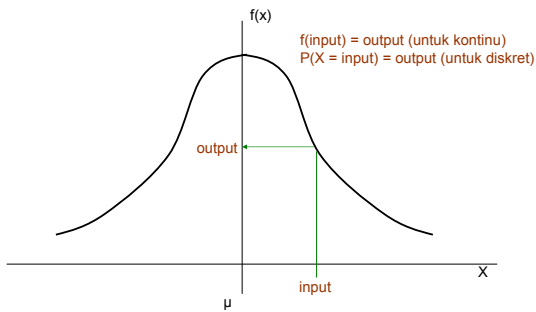
$$z = \frac{X - \mu}{\sigma}$$

- MINITAB: Calc -> Probability Distribution -> Normal

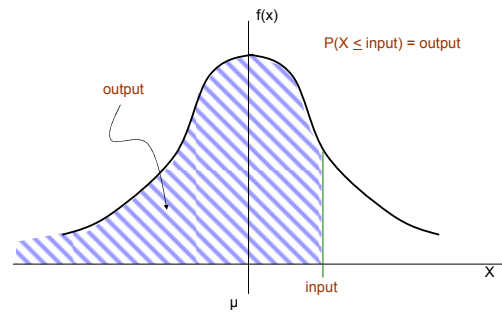
Pilihan Distribusi Probabilitas di dalam MINTAB

- Calc -> Probability Distribution -> [nama distribusinya, misalnya Normal].
- Ada 3 Pilihan:
 - Probability Density
 - Cumulative Probability
 - Inverse Cumulative Probability

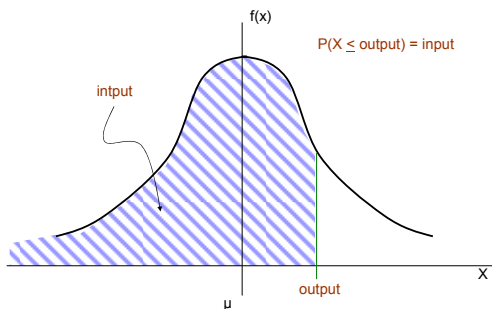
Probability Density



Cumulative Probability

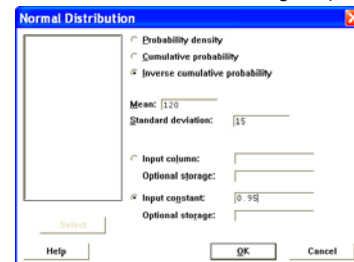


Inverse Cumulative Probability



Contoh Soal Distribusi Kontinu

- Contoh: Diketahui X terdistribusi normal dengan rata-rata 120 dan deviasi standar 15. Carilah x agar $P(X > x) = 5\%$.



- Ans: $x = 144.6728$

Pendekatan Normal untuk Binomial

- Binomial: diskret, parameter n dan p
- Normal: kontinu, parameter μ dan σ
- Untuk n besar, distribusi binomial akan menyerupai distribusi normal. Jadi untuk masalah binomial dengan n besar, dapat didekati dengan distribusi normal
- Ingat:
 - Untuk diskret: $P(X=x) = \text{ada nilai}$
 - Untuk kontinu: $P(X=x) = 0$

Contoh Pendekatan Normal untuk Binomial

- Untuk X yang terdistribusi binomial dengan $n = 80$ dan $p = 0.3$, carilah
 - $P(X=24)$
 - $P(X>30)$
 - $P(30 < X \leq 34)$
 - $P(X \leq 33)$

Jawab:

- Untuk distribusi binomial:
 - Rata-rata = $\mu = np = 80 \cdot 0.3 = 24$
 - Deviasi Standar = $\sigma = \sqrt{n \cdot p \cdot q} = 4.0988$
 - Rata-rata dan deviasi standar tersebut digunakan sebagai parameter distribusi normal
- $$P(X = 24) = P(23.5 < X < 24.5) = P\left(\frac{23.5 - 24}{4.0988} < Z < \frac{24.5 - 24}{4.0988}\right)$$
- ↑ diskret
↑ kontinu
↑ koreksi kontinuitas
- $$= P(-0.122 < Z < 0.122) = 2 \cdot 0.0478 = 0.0956$$

- Cek dengan rumus Binomial:

$$P(X = 24) = \frac{80!}{24!(80-24)!} 0.3^{24} 0.7^{80-24} = 0.0969513$$

$$P(X > 30) = P(X > 30.5) = P\left(Z > \frac{30.5 - 24}{4.0988}\right)$$

↑ diskret
↑ kontinu
↑ koreksi kontinuitas

$$= P(Z > 1.5858) = 0.5 - 0.4441 = 0.0559$$

$$P(30 < X \leq 34) = P(30.5 < X < 34.5) = P\left(\frac{30.5 - 24}{4.0988} < Z < \frac{34.5 - 24}{4.0988}\right)$$

↑ diskret
↑ kontinu
↑ koreksi kontinuitas

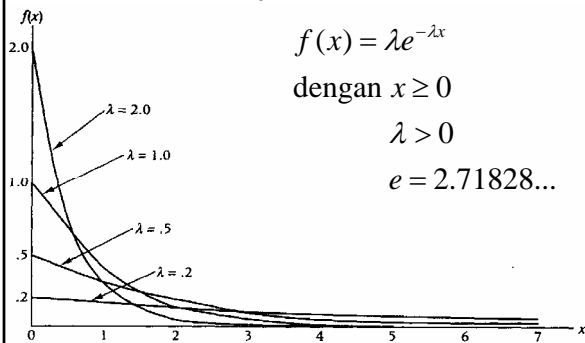
$$= P(1.5858 < Z < 2.5617) = 0.4948 - 0.4441 = 0.0507$$

$$P(X \leq 33) = P(X < 33.5) = P\left(Z < \frac{33.5 - 24}{4.0988}\right)$$

↑ diskret
↑ kontinu
↑ koreksi kontinuitas

$$= P(Z < 2.3177) = 0.5 + 0.4898 = 0.9898$$

Distribusi Eksponensial

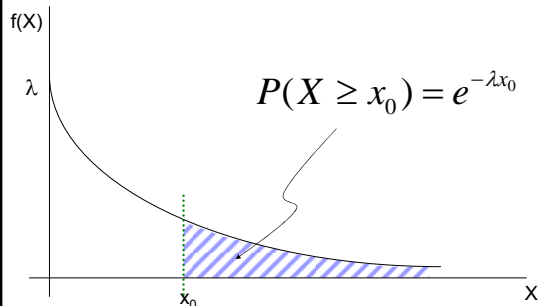


Distribusi Eksponensial (lanjutan)

- Adalah distribusi kontinu
- Adalah kelompok distribusi dengan parameter $= \lambda$, yang terjadi pada $X = 0$
- Mempunyai ekor di sebelah kanan
- Nilai x mulai dari nol sampai tak hingga
- Puncaknya selalu ada di $X = 0$
- Kurvanya selalu mengecil untuk X yang membesar
- Menunjukkan distribusi probabilitas untuk waktu antara kejadian acak
- Rata-rata dan deviasi standarnya:

$$\mu = \frac{1}{\lambda} \quad \text{dan} \quad \sigma = \frac{1}{\lambda}$$

Distribusi Eksponensial (lanjutan)



Contoh Soal Distribusi Eksponensial

- Di restoran sebuah kota kecil kedatangan pelanggan dapat dianggap terdistribusi Poisson dengan rata-rata 3.2 pelanggan per 30 menit.
 - Berapa menit waktu rata-rata antar kedatangan pelanggan di restoran tersebut?
 - Berapa probabilitas bahwa antar kedatangan pelanggan ada selang 1 jam atau kurang?
 - Berapa probabilitas bahwa dua pelanggan datang dengan selang waktu kedatangan 15 menit atau lebih?

Jawab

- $\mu = 1/3.2 = 0.313$. Jadi rata-rata $0.313 \cdot 30$ menit = 9.39 menit waktu antar kedatangan pelanggan
- 1 jam = 2 interval, yaitu $2 \cdot 30$ menit. Jadi $x = 2$. $P(X \geq 2) = 1 - \exp(-3.2 \cdot 2) = \underline{0.998}$
- 15 menit = 0.5 interval. Jadi $x = 0.5$. $P(X \geq 0.5) = \exp(-3.2 \cdot 0.5) = \underline{0.202}$

Bagian 5

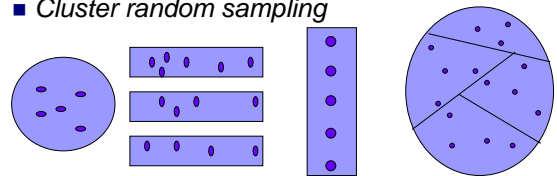
Sampling dan Distribusi Sampling

Sampling (pengambilan sampel)

- Dapat menghemat biaya
- Dapat menghemat waktu
- Untuk sumberdaya yang terbatas, pengambilan sampel dapat memperluas cakupan studi
- Bila proses riset bersifat destruktif, pengambilan sampel dapat menghemat produk
- Apabila akses ke seluruh populasi tidak dapat dilakukan, pengambilan sampel adalah satu-satunya pilihan

Random Sampling

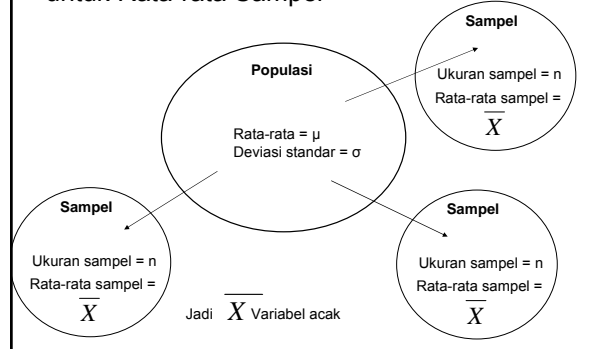
- *Simple random sampling*
- *Stratified random sampling*
- *Systematic random sampling*
- *Cluster random sampling*



Nonrandom Sampling

- Convenience sampling
- Judgement sampling
- Quota sampling

Sampling Distribution (distribusi sampling) untuk Rata-rata Sampel

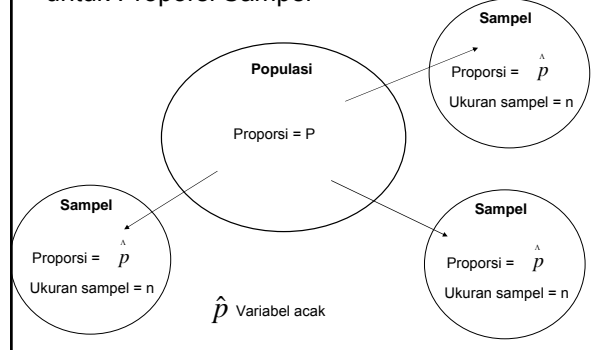


Teorema Limit Tengah untuk Rata-rata

- Apabila sampel berukuran n besar (≥ 30) diambil dari populasi yang mempunyai rata-rata μ dan deviasi standar σ , maka rata-rata sampel \bar{X} akan terdistribusi normal dengan rata-rata μ dan deviasi standar σ/\sqrt{n}
- Khusus: apabila populasinya terdistribusi normal, maka n pada teorema di atas tidak harus besar. Jadi

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{adalah normal standar}$$

Sampling Distribution (distribusi sampling) untuk Proporsi Sampel



Teorema Limit Tengah untuk Proporsi

- Apabila sampel berukuran n diambil dari populasi yang proporsinya P, dengan $n \cdot P > 5$ dan $n \cdot Q > 5$, maka proporsi sampel \hat{p} akan terdistribusi normal dengan rata-rata P dan deviasi standar $\sqrt{(P \cdot Q/n)}$. Jadi

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{P \cdot Q}{n}}} \quad \text{adalah Normal standar}$$

Bagian 6

Estimasi untuk Populasi Tunggal

Statistika Inferensial

Populasi

Sampel

Simpulkan (estimasi) tentang parameter

mendapatkan statistik

- Estimasi: Estimasi titik, estimasi interval, uji hipotesa

Estimasi Interval untuk μ

- Selang kepercayaan 100(1- α)% untuk μ pada sampel besar: $\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- Artinya: $P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 100(1-\alpha)\%$

Distribusi Normal Standar

$\frac{\alpha}{2}$

$1-\alpha$

0

$-\frac{Z_{\frac{\alpha}{2}}}{2}$

$\frac{Z_{\frac{\alpha}{2}}}{2}$

Z

Note: apabila σ tidak diketahui dapat digantikan dengan s

Estimasi Interval untuk μ (lanjutan)

- MINITAB: Stat -> Basic Statistics -> 1-sample z

1-Sample z

Variables: HrgTanah

Confidence Interval Level: 95.0

Test mean: 0.0

Alternative: not equal

Sigma: 120

Estimasi Interval untuk μ (lanjutan)

Output MINITAB:

Confidence Interval s

The assumed sigma = 120

Variable	N	Mean	StDev	SE Mean	95.0 % CI
HrgTanah	50	924.2	174.7	17.0	(890.9, 957.5)

Estimasi Interval untuk μ , sampel kecil.

Asumsi: Populasi terdistribusi Normal

- Selang kepercayaan 100(1- α)% untuk μ pada sampel kecil: $\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$
- Artinya: $P\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}\right) = 100(1-\alpha)\%$

distribusi t dengan df = n-1

$\frac{\alpha}{2}$

$1-\alpha$

0

$-\frac{t_{\frac{\alpha}{2}, n-1}}{2}$

$\frac{t_{\frac{\alpha}{2}, n-1}}{2}$

t

Estimasi Interval untuk μ sampel kecil (lanjutan)

- MINITAB: Stat -> Basic Statistics -> 1-sample t

1-Sample t

Variables: HrgTanah

Confidence Interval Level: 95.0

Test mean: 0.0

Alternative: not equal

Estimasi Interval untuk μ sampel kecil (lanjutan)

Output MINITAB:

Confidence Intervals

Variable	N	Mean	StDev	SE Mean	95.0 % CI
HrgTanah	15	952.7	243.4	62.8	(817.9, 1087.5)

Estimasi Interval untuk P

Syarat: $nP > 5$ dan $nQ > 5$

Populasi

Proporsi = P
(akan diestimasi)

→

Sampel

Ukuran = n
proporsi \hat{p}

$\hat{q} = 1 - \hat{p}$

- Selang kepercayaan 100(1- α)% untuk P $\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$
- Artinya: $P\left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq P \leq \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 100(1-\alpha)\%$

Estimasi Interval untuk Varians Populasi σ^2

- Syarat:
 - Populasi terdistribusi Normal
 - Sampel besar

Populasi

Varians Populasi = σ^2
(akan diestimasi)

→

Sampel

Ukuran = n
Varians sampel = S^2

Estimasi Interval untuk Varians Populasi σ^2

- Selang kepercayaan 100(1- α)% untuk varians populasi σ^2

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right)$$

- Artinya $P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right) = 100(1-\alpha)\%$

Estimasi Interval untuk Varians Populasi σ^2 (lanjutan)

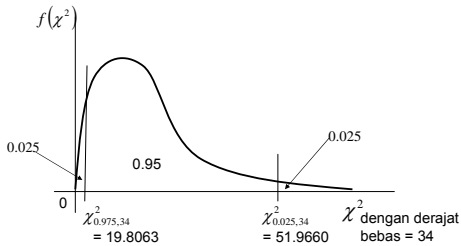
- Distribusi χ^2

dengan derajat bebas = n-1

Estimasi Interval untuk Varians Populasi σ^2 (lanjutan)

- Contoh distribusi χ^2 untuk df = 34 dan $\alpha = 0.05$
- MINITAB: Calc -> Probability Distribution -> Chisquare

Estimasi Interval untuk Varians Populasi σ^2 (lanjutan)



Ukuran Sampel dalam Mengestimasi Rata-rata Populasi μ

- Dalam mengestimasi rata-rata populasi μ , ukuran sampel minimum untuk suatu α dan E yang ditetapkan, adalah

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

dengan

E = galat estimasi = *error of estimation* = $\bar{x} - \mu$

σ = deviasi standar populasi,

= *range/4* apabila tidak diketahui

α = taraf keterandalan

$100(1 - \alpha)\%$ = tingkat keyakinan = *level of confidence*

Contoh

- Seorang manajer bank ingin menentukan rata-rata deposito bulanan per nasabah di bank tersebut. Untuk itu ia akan mengestimasi dengan menggunakan selang kepercayaan. Berapa ukuran sampel yang harus ia ambil apabila ia ingin yakin 99% dan kesalahannya tidak lebih dari 200 juta rupiah. Ia asumsikan bahwa deviasi standar untuk deposito bulanan semua nasabah adalah 1 milyar rupiah

Jawab

- X = besarnya deposito bulanan nasabah, dinyatakan dalam juta rupiah
- $\sigma = 1000$
- Tingkat keyakinan 99% $\rightarrow \alpha = 0.01$ dan $\alpha/2 = 0.005$, sehingga $z_{\alpha/2} = z_{0.005} = 2.5758$
- $E = 200$
- Ukuran sampel minimum

$$n = \left(\frac{z_{\alpha/2} \sigma}{E} \right)^2 = \left(\frac{2.5758 * 1000}{200} \right)^2 = 165.87 = 166$$

Ukuran Sampel dalam Mengestimasi Proporsi Populasi p

- Dalam mengestimasi proporsi populasi p , ukuran sampel minimum untuk suatu α dan E yang ditetapkan, adalah

$$n = \frac{z_{\alpha/2}^2 pq}{E^2}$$

dengan

E = galat estimasi = *error of estimation* = $\hat{p} - p$

p = proporsi populasi,

= 0.5 apabila tidak diketahui (agar n maksimum)

$q = 1 - p$

α = taraf keterandalan

$100(1 - \alpha)\%$ = tingkat keyakinan = *level of confidence*

Contoh

- Seseorang ingin menyelidiki berapa proporsi sekretaris di seluruh perkantoran di Bandung yang dilengkapi dengan komputer di ruang kerjanya. Ia akan menjawab pertanyaan ini dengan melakukan survei acak. Berapa ukuran sampel yang harus ia ambil apabila ia ingin yakin 95% dan galat pada selang kepercayaan tidak dapat lebih dari 0.05? Anggap bahwa proporsi aktual tidak diketahui sebelumnya.

Jawab

- p = proporsi sekretaris di seluruh perkantoran di Bandung yang dilengkapi dengan komputer di ruang kerjanya
- Karena p tidak diketahui, asumsikan nilainya 0.5
- q = 1 - p = 0.5
- Tingkat keyakinan 95% → $\alpha = 0.05$ dan $\alpha/2 = 0.025$, sehingga $z_{\alpha/2} = z_{0.025} = 1.96$
- E = 0.05
- Ukuran sampel minimum

$$n = \frac{z_{\alpha/2}^2 pq}{E^2} = \frac{1.96^2 * 0.5 * 0.5}{0.05^2} = 384.16 = 385$$

Bagian 7

Uji Hipotesa untuk Populasi Tunggal

Uji Hipotesis

- Hipotesis Riset: menyatakan hubungan
 - Hipotesa nol (H_0) vs Hipotesa alternatif ($H_1 = H_a$)
 - Galat (*error*) tipe I, galat tipe II, dan *power*

	H_0 benar	H_0 salah
Pertahankan H_0	Keputusan benar	Galat Tipe II (β)
Tolak H_0	Galat Tipe I (α)	Keputusan benar (<i>power</i>)

- R = *Rejection Region*. Apabila statistik uji (test statistic) ada di daerah ini, maka tolak H_0 . Bila tidak, maka pertahankan H_0 .

Beberapa Uji Hipotesis pada Statistika Parametrik

- Uji z 1 sampel: mengestimasi rata-rata populasi dengan menggunakan sampel besar
- Uji t 1 sampel: mengestimasi rata-rata populasi dengan menggunakan sampel kecil pada populasi yang terdistribusi normal
- Uji t 2 sampel: mengestimasi perbedaan rata-rata 2 populasi independen dengan menggunakan sampel kecil pada populasi yang terdistribusi normal
- Anova 1 arah (*completely randomized design*): mempelajari apakah rata-rata c populasi semuanya sama, atau ada yang berbeda
- Anova 2 arah (*factorial design*):
 - mempelajari apakah rata-rata c populasi semuanya sama, atau ada yang berbeda
 - mempelajari apakah rata-rata r populasi semuanya sama, atau ada yang berbeda
 - mempelajari apakah efek interaksi ada atau tidak ada

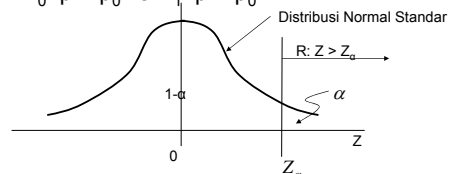
Beberapa Uji Hipotesis pada Statistika Nonparametrik

- Uji U Mann-Whitney: membandingkan dua populasi independen
- Uji peringkat bertanda Wilcoxon: membandingkan dua populasi yang *related*
- Uji K Kruskal-Wallis: menguji apakah c populasi identik atau berbeda pada *completely random design*
- Uji Friedman: menguji apakah c populasi identik atau berbeda, pada *randomized block design*

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Besar, Uji z 1 sampel

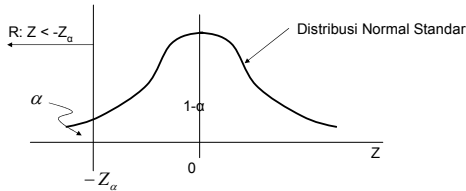
■ Statistik uji
$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$



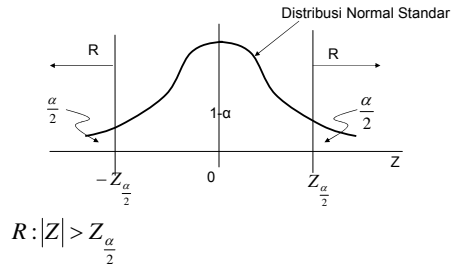
Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Besar, Uji z 1 sampel (lanjutan)

- $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$



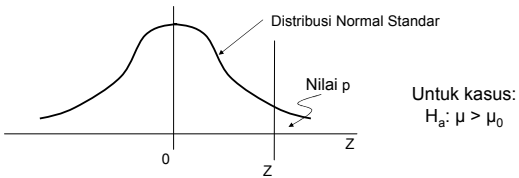
Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Besar, Uji z 1 sampel (lanjutan)

- $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

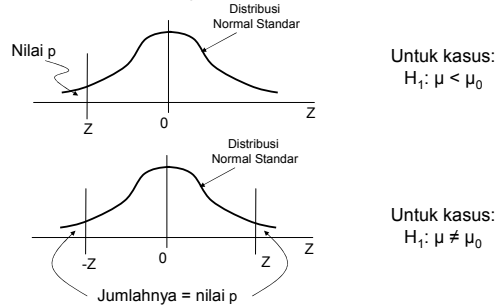


Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Besar, Uji z 1 sampel (lanjutan)

- Cara lain: dengan menggunakan nilai p (*p-value*), berlaku untuk ketiga hipotesa alternatif:
Tolak H_0 jika $p < \alpha$



Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Besar, Uji z 1 sampel (lanjutan)

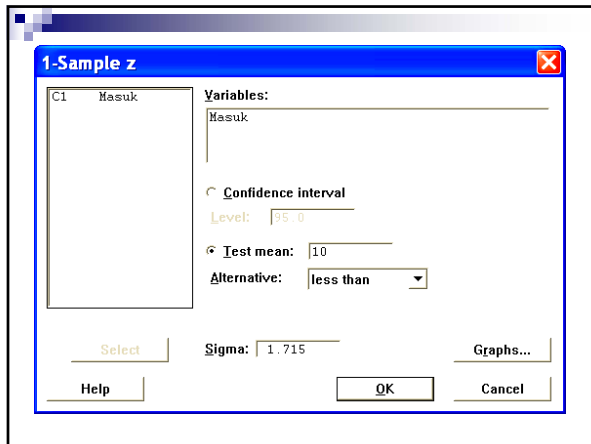


Contoh Aplikasi Uji Z 1 sampel

- Sebuah laporan menyebutkan bahwa rata-rata penjualan harian di restoran A tidak melebihi 10 juta rupiah. Untuk menguji apakah hal ini benar, maka dikumpulkan data penjualan di restoran A selama 30 hari (dalam juta rupiah). Gunakanlah taraf keterandalan $\alpha = 5\%$. Kesimpulan apakah yang dapat ditarik?

Data	
9.7	10.5
8.5	10.2
9.8	5.5
11.0	7.0
11.5	7.2
13.0	8.0
8.7	8.0
7.9	9.5
8.4	9.5
7.6	7.8
10.6	10.5
10.9	11.0
11.0	12.0
9.1	9.8
10.0	7.0

MINITAB: Stat -> Basic Statistics -> 1-sample Z
Note: Hitung dahulu deviasi standar sampel, S



Output MINITAB

Z-Test

Test of mu = 10.000 vs mu < 10.000
The assumed sigma = 1.71

Variable	N	Mean	StDev	SE Mean	Z	P
Masuk	30	9.373	1.715	0.313	-2.00	0.023

- Dengan metode nilai p: terlihat bahwa nilai p = 0.023 < α = 0.05. Jadi, tolak H₀. Artinya: rata-rata penjualan di restoran A tidak melebihi 10 juta rupiah.
- Dengan metode nilai kritis: Z = -2.00 berada di R, yaitu Z < -1.645. Kesimpulan: tolak H₀. Artinya: rata-rata penjualan di restoran A tidak melebihi 10 juta rupiah.

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Kecil, Uji t 1 sampel. Asumsi: Populasi Terdistribusi Normal

- Statistik uji $t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$
- H₀: μ = μ₀ vs H₁: μ > μ₀

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Kecil, Uji t 1 sampel. Asumsi: Populasi Terdistribusi Normal (lanjutan)

- H₀: μ = μ₀ vs H₁: μ < μ₀

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Kecil, Uji t 1 sampel. Asumsi: Populasi Terdistribusi Normal (lanjutan)

- H₀: μ = μ₀ vs H₁: μ ≠ μ₀

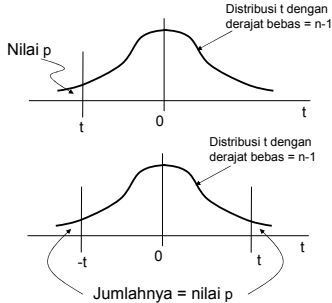
$R: |t| > t_{\frac{\alpha}{2}, n-1}$

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Kecil, Uji t 1 sampel. Asumsi: Populasi Terdistribusi Normal (lanjutan)

- Cara lain: dengan menggunakan nilai p (*p-value*), berlaku untuk ketiga hipotesa alternatif:
Tolak H₀ jika p < α

Untuk kasus: H₁: μ > μ₀

Uji Hipotesis tentang Rata-rata Populasi dengan menggunakan Sampel Kecil, Uji t 1 sampel.
Asumsi: Populasi Terdistribusi Normal (lanjutan)



Untuk kasus:
 $H_1: \mu < \mu_0$

Untuk kasus:
 $H_1: \mu \neq \mu_0$

Contoh Aplikasi Uji t 1 sampel

- Majalah A menyebutkan bahwa rata-rata usia direktur utama bank di sebuah kota 41 tahun. Untuk menguji apakah hal ini benar, maka dikumpulkanlah data acak dari 11 direktur utama bank di kota tersebut. Asumsikan bahwa usia direktur utama bank di kota tersebut terdistribusi normal. Gunakanlah taraf keterandalan $\alpha = 5\%$. Kesimpulan apakah yang dapat ditarik?
- Data: 40, 43, 44, 50, 39, 38, 51, 37, 55, 57, 41
- MINITAB: Stat -> Basic Statistics -> 1 Sample t

1-Sample t

Variables:
Usia

Confidence interval
Level: 95.0

Test mean: 41
Alternative: not equal

Select OK Cancel

Output MINITAB

T-Test of the Mean

Test of $\mu = 41.00$ vs $\mu \text{ not } = 41.00$

Variabl e	N	Mean	StDev	SE Mean	T	P
Usi a	11	45.00	7.07	2.13	1.88	0.090

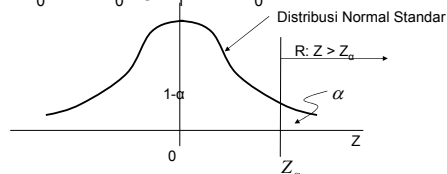
• Dengan metode nilai p: terlihat bahwa nilai $p = 0.090 > \alpha = 0.05$. Jadi, pertahankan H_0 . Artinya: data yang ada mendukung pernyataan bahwa rata-rata usia direktur bank di kota tersebut 41 tahun.

• Dengan metode nilai kritis: $t = 1.88$ berada di luar R, yaitu $|t| < 2.2281$. Kesimpulan: pertahankan H_0 (sama dengan kesimpulan di atas).

Uji Hipotesis tentang Proporsi Populasi.
 $nP > 5$ dan $nQ > 5$

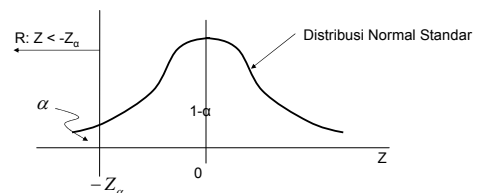
■ Statistik uji
$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0 Q_0}{n}}}$$

■ $H_0: P = P_0$ vs $H_1: P > P_0$



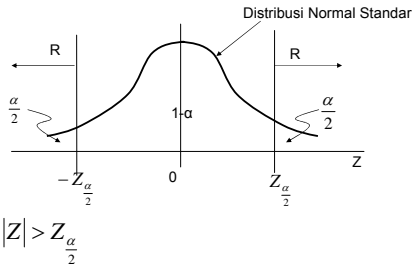
Uji Hipotesis tentang Proporsi Populasi.
 $nP > 5$ dan $nQ > 5$ (lanjutan)

■ $H_0: P = P_0$ vs $H_1: P < P_0$



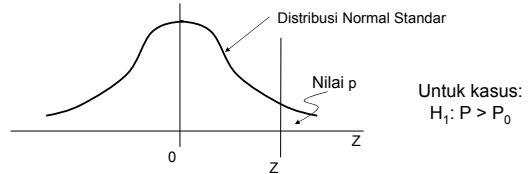
**Uji Hipotesis tentang Proporsi Populasi.
nP>5 dan nQ>5 (lanjutan)**

- $H_0: P = P_0$ vs $H_1: P \neq P_0$

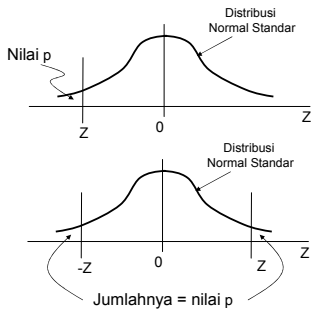


**Uji Hipotesis tentang Proporsi Populasi.
nP>5 dan nQ>5 (lanjutan)**

- Cara lain: dengan menggunakan nilai p (*p-value*), berlaku untuk ketiga hipotesa alternatif:
Tolak H_0 jika $p < \alpha$



**Uji Hipotesis tentang Proporsi Populasi.
nP>5 dan nQ>5 (lanjutan)**



**Contoh Uji Hipotesis Tentang
Proporsi Populasi**

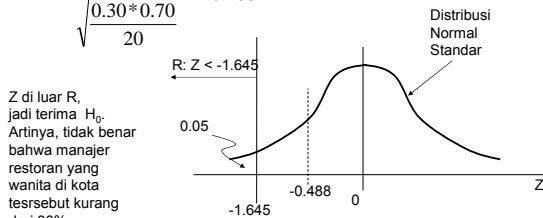
- Untuk menyelidiki kebenaran apakah manajer restoran yang wanita di sebuah kota kurang dari 30%, seseorang mengumpulkan data dari 20 restoran di kota tersebut yang diambil secara acak. Hasilnya: ada 5 restoran yang manajernya wanita, sisanya mempunyai manajer pria. Apa kesimpulan dari data tersebut, apabila α yang digunakan 5%?

Jawab

- $H_0: P = 0.30$ vs $H_1: P < 0.30$

$$\hat{p} = \frac{5}{20} = 0.25$$

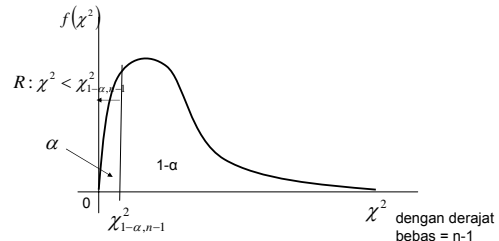
$$Z = \frac{0.25 - 0.30}{\sqrt{\frac{0.30 \cdot 0.70}{20}}} = -0.488$$



Z di luar R, jadi terima H_0 . Artinya, tidak benar bahwa manajer restoran yang wanita di kota tersebut kurang dari 30%

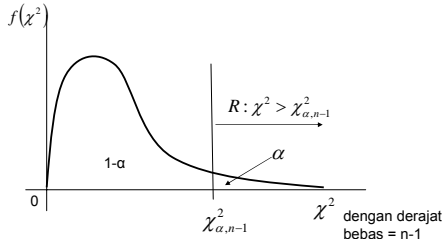
**Uji Hipotesis tentang Varians Populasi.
Asumsi: Populasi Terdistribusi Normal**

- Statistik uji $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$
- $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$



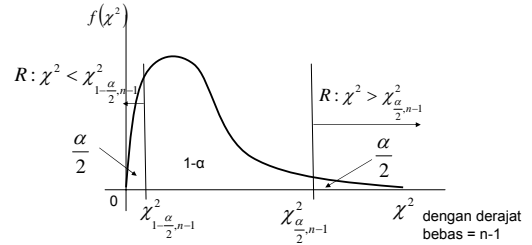
Uji Hipotesis tentang Varians Populasi.
Asumsi: Populasi Terdistribusi Normal (lanjutan)

■ $H_0: \sigma^2 = \sigma^2_0$ vs $H_1: \sigma^2 > \sigma^2_0$



Uji Hipotesis tentang Varians Populasi.
Asumsi: Populasi Terdistribusi Normal (lanjutan)

■ $H_0: \sigma^2 = \sigma^2_0$ vs $H_1: \sigma^2 \neq \sigma^2_0$



Contoh Uji Hipotesis Tentang Varians Populasi

- Spesifikasi mesin pemotong menyebutkan bahwa deviasi standar hasil potongan kurang dari 6 mm. Untuk menguji hal ini, dikumpulkan 30 hasil potongan mesin tersebut. Dengan menggunakan $\alpha = 10\%$, kesimpulan apakah yang dapat ditarik dari data tersebut.

Data

105	100
110	101
106	102
111	103
101	100
100	103
101	99
101	99
100	98
95	98
97	94
108	100
99	100
99	101
100	100

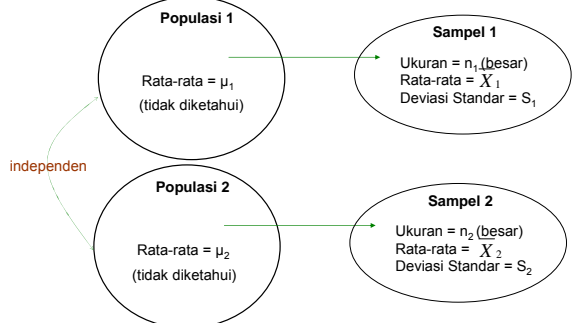
Jawab

- $N = 30$
- $S = 3.82$ (Stat -> Basic Statistic -> Descriptive Statistics)
- $H_0: \sigma^2 = 36$ vs $H_1: \sigma^2 < 36$
- Untuk $df = 29$ dan $\alpha = 0.10$, $\chi^2_{0.90, 29} = 19.7677$ (Calc -> Probability Distribution -> Chisquare)
- $R: \chi^2 < \chi^2_{0.90, 29} = 19.7677$
- Statistik uji: $\chi^2 = \frac{(30-1)3.82^2}{6^2} = 11.7550$
- Karena $11.7550 < 19.7677$, maka tolak H_0 . Artinya, benar bahwa deviasi standar hasil potong mesin tersebut kurang dari 6 mm.

Bagian 8

Statistika Inferensi untuk Dua Populasi

Statistika Inferensi Tentang Rata-rata Dua Populasi Independen

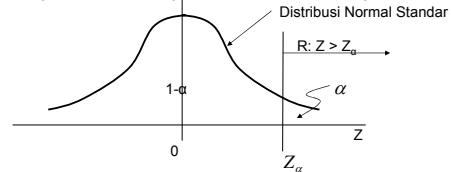


Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Besar

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

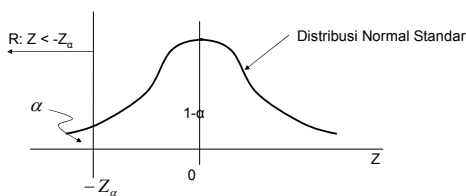
Catatan: Bila deviasi standar populasi σ tidak ada, dapat digantikan dengan deviasi standar sampel S

$H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 > \mu_0$



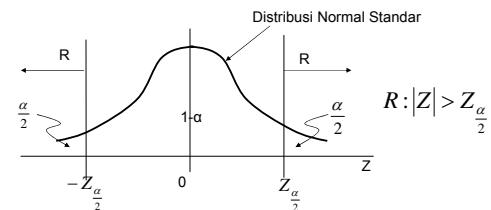
Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Besar (lanjutan)

$H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 < \mu_0$



Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Besar (lanjutan)

$H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 \neq \mu_0$



Catatan: sebagai alternatif, metode nilai p juga dapat digunakan

Selang Kepercayaan 100(1-α)% Perbedaan Rata-rata 2 Populasi Independen $\mu_1 - \mu_2$ dengan menggunakan Sampel Besar

$$\left[(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

Artinya:

$$P \left((\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 100(1-\alpha)\%$$

Catatan: Bila deviasi standar populasi σ tidak ada, dapat digantikan dengan deviasi standar sampel S

Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Kecil. Asumsi: Kedua Populasi terdistribusi Normal dan Deviasi standar kedua populasi sama

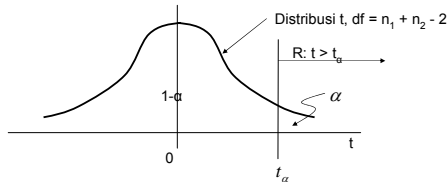
Statistik uji $t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$S =$ pooled standard deviation

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

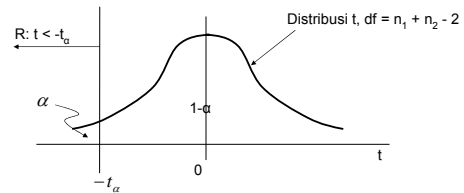
Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Kecil. Asumsi: Kedua Populasi terdistribusi Normal dan Deviasi standar kedua populasi sama (lanjutan)

■ $H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 > \mu_0$



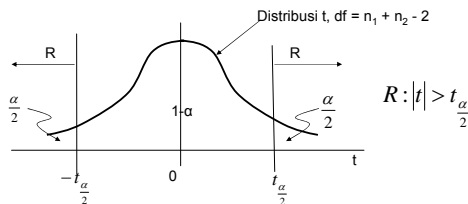
Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Kecil. Asumsi: Kedua Populasi terdistribusi Normal dan Deviasi standar kedua populasi sama (lanjutan)

■ $H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 < \mu_0$



Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Independen dengan menggunakan Sampel Kecil. Asumsi: Kedua Populasi terdistribusi Normal dan Deviasi standar kedua populasi sama (lanjutan)

■ $H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_1: \mu_1 - \mu_2 \neq \mu_0$



■ Catatan: sebagai alternatif, metode nilai p juga dapat digunakan

Selang Kepercayaan 100(1-alpha)% Perbedaan Rata-rata 2 Populasi Independen $\mu_1 - \mu_2$ dengan menggunakan Sampel Kecil. Asumsi: Kedua populasi terdistribusi normal dan deviasi standarnya sama

$$\left(\bar{X}_1 - \bar{X}_2 \right) \pm t_{\frac{\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Derajat bebas t adalah $n_1 + n_2 - 2$

Artinya:

$$P\left((\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) = 100(1 - \alpha)\%$$

Contoh Uji Hipotesis Perbedaan Rata-rata 2 Populasi dengan menggunakan Sampel Kecil

■ Sebuah laporan menyebutkan bahwa rata-rata gaji bulanan direktur bank di Jakarta lebih tinggi dari pada di Bandung. Untuk menyelidiki kebenaran hal ini, seorang peneliti mengumpulkan data yang diambil secara acak di Jakarta dan di Bandung, sebagaimana tercantum dalam data berikut (dalam juta rupiah). Dengan menggunakan taraf keterandalan $\alpha = 5\%$, kesimpulan apa yang dapat ditarik mengenai laporan tersebut di atas.

Data

Row	Jakarta	Bandung
1	5.6	8.1
2	7.1	7.9
3	6.8	5.4
4	10.2	4.5
5	12.5	5.6
6	13.5	6.8
7	6.8	9.2
8	5.8	8.1
9	9.9	7.2
10	10.2	4.5
11	15.6	5.2
12	7.7	6.8
13	9.8	6.7
14	6.8	5.7
15	5.8	5.8
16	6.8	5.8
17	8.9	10.3
18	9.4	4.5
19	10.5	5.8
20	12.6	10.2
21		9.8
22		5.8
23		5.5
24		5.6
25		7.2

Solusi (asumsi: gaji bulanan direktur bank di Bandung dan Jakarta terdistribusi normal)

$H_0: \mu_j - \mu_b = 0$ vs $H_1: \mu_j - \mu_b > 0$

Two-sample T for j vs b

	N	Mean	StDev	SE Mean
j	20	9.12	2.83	0.63
b	25	6.72	1.75	0.35

Difference = mu j - mu b

Estimate for difference: 2.395

95% lower bound for difference: 1.240

T-Test of difference = 0 (vs >): T-Value = 3.48 P-Value = 0.001 DF = 43

Both use Pooled StDev = 2.29

Kesimpulan: tolak $H_0: \mu_j - \mu_b = 0$. Jadi: laporan bahwa rata-rata gaji bulanan direktur bank di Jakarta lebih tinggi dari pada di Bandung didukung data.

Solusi (asumsi: gaji bulanan direktur bank di Bandung dan Jakarta tidak terdistribusi normal) -> Statistika Nonparametrik

$H_0: \mu_j - \mu_b = 0$ vs $H_1: \mu_j - \mu_b > 0$

Mann-Whitney Test and CI: Jakarta, Bandung

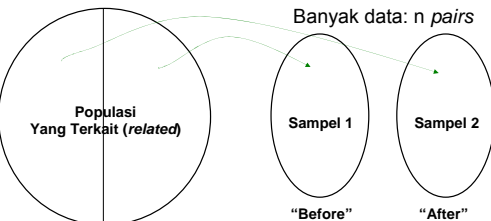
Jakarta	N = 20	Median = 9.150
Bandung	N = 25	Median = 5.800
Point estimate for ETA1-ETA2 is	2.100	
95.2 Percent CI for ETA1-ETA2 is	(0.899, 3.800)	
W =	593.5	

Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0012

The test is significant at 0.0012 (adjusted for ties)

Kesimpulan: tolak $H_0: \mu_j - \mu_b = 0$. Jadi: laporan bahwa rata-rata gaji bulanan direktur bank di Jakarta lebih tinggi dari pada di Bandung didukung data.

Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Terkait (*related*) dengan menggunakan Sampel Kecil. Asumsi: Perbedaan tersebut Terdistribusi Normal



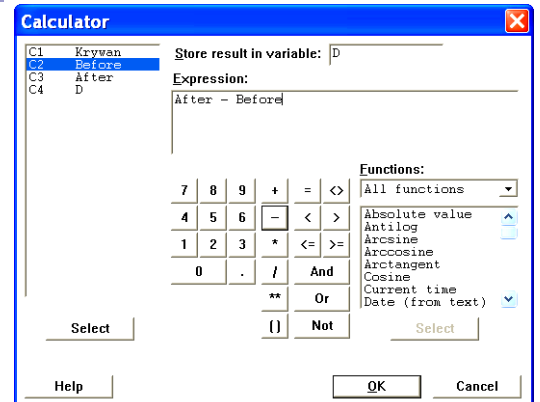
- Hitung d = perbedaan antara *before* dan *after* untuk setiap pasang data. Selanjutnya, lakukan uji t 1 sampel dengan data d tersebut.

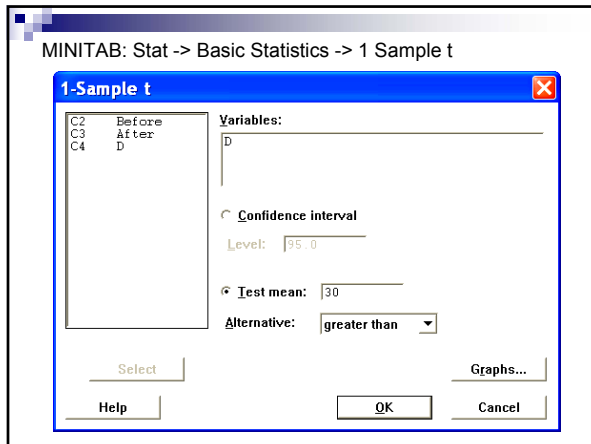
Contoh Aplikasi. Uji Hipotesis tentang Perbedaan Rata-rata 2 Populasi Terkait (*related*) dengan menggunakan Sampel Kecil

- Sebuah lembaga kursus Bahasa Inggris mengklaim bahwa apabila seseorang mengikuti kursus selama 2 bulan di lembaga tersebut, maka nilai TOEFL orang tersebut akan meningkat sedikitnya 30. Untuk menguji klaim tersebut, 11 orang diukur nilai TOEFL mereka sebelum dan sesudah mengikuti kursus Bahasa Inggris di lembaga tersebut. Data terlampir. Dengan menggunakan $\alpha = 10\%$, kesimpulan apakah yang dapat ditarik mengenai klaim lembaga tersebut? Asumsikan perbedaan nilai TOEFL sebelum dan sesudah kursus terdistribusi normal

Data	Row	Karyawan	Before	After	D
	1	Adi	450	470	20
	2	Budi	503	535	32
	3	Ci ca	400	433	33
	4	Dedi	435	450	15
	5	Edi	370	450	80
	6	Feri	550	570	20
	7	Gi na	525	555	30
	8	Hedi	378	410	32
	9	I wan	440	480	40
	10	Joni	510	555	45
	11	Ki a	522	535	13
	12	Lena	533	566	33

D = after - before
Untuk menghitung D, Calc -> Calculator





Output MINITAB

T-Test of the Mean

Test of mu = 30.00 vs mu > 30.00

Variable	N	Mean	StDev	SE Mean	T	P
D	12	32.75	17.77	5.13	0.54	0.30

Nilai p = 0.30 dan $\alpha = 0.10$. Ternyata nilai $p > \alpha$, maka terima H_0 .
Kesimpulan: klaim lembaga kursus Bahasa Inggris bahwa setelah kursus peningkatan nilai TOEFL sedikitnya 30, **tidak didukung data**.

Estimasi Interval untuk d, sampel kecil.

Asumsi: Populasi terdistribusi Normal

- Selang kepercayaan 100(1- α)% untuk d pada sampel kecil: $\bar{d} \pm t_{\frac{\alpha}{2}, n-1} \frac{S_d}{\sqrt{n}}$
- Artinya: $P\left(-t_{\frac{\alpha}{2}, n-1} \frac{S_d}{\sqrt{n}} \leq d \leq \bar{d} + t_{\frac{\alpha}{2}, n-1} \frac{S_d}{\sqrt{n}}\right) = 100(1-\alpha)\%$

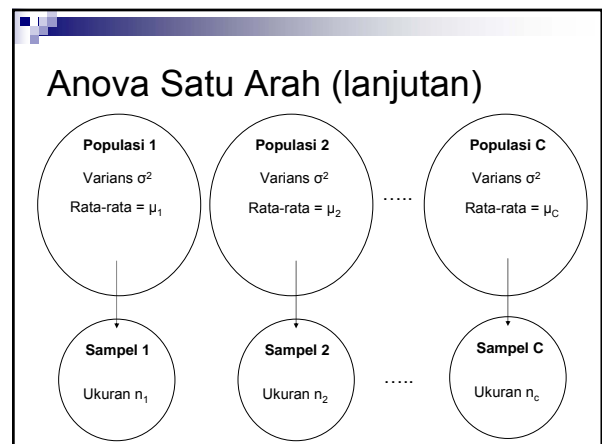
distribusi t dengan df = n-1

Bagian 9

Anova

Anova Satu Arah (One Way Anova)

- Membandingkan C (>2) populasi independen (*completely randomized design*)
- Asumsi:
 - Populasi terdistribusi normal
 - Sampel diambil secara acak dari masing-masing populasi
 - Varians semua populasi sama
- $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_C$
- H_1 : sedikitnya ada 1 rata-rata populasi yang berbeda



Anova Satu Arah (lanjutan)

Source	DF	SS	MS	F
Treatment (C = Column)	C - 1	SSC	$MSC = \frac{SSC}{C-1}$	$F = \frac{MSC}{MSE}$
Error	N - C	SSE	$MSE = \frac{SSE}{N-C}$	
Jumlah	N - 1	SST		

Catatan: $N = \sum_{i=1}^c n_i$
 Derajat bebas F adalah C-1 (pembilang) dan N-C (penyebut)

Contoh Aplikasi Anova Satu Arah

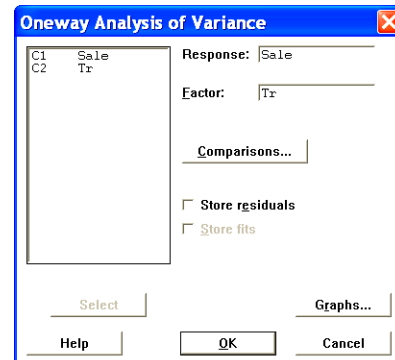
- Untuk mengetahui apakah ada pengaruh kemasan suatu produk kecantikan terhadap penjualannya, sebuah pabrik alat-alat kecantikan melakukan pengujian dengan membuat 4 macam kemasan, yaitu A, B, C, D. Penjualan selama beberapa bulan (dalam juta rupiah) untuk masing-masing kemasan dicatat (terlampir). Dengan menggunakan $\alpha = 5\%$, kesimpulan apakah yang dapat ditarik?

Data

A	B	C	D
15	8	11	14
11	8	11	11
10	7	8	10
9	9	8	9
	11	9	11
	8	10	12
			12
			10

Row	Sale	Tr
1	15	1
2	11	1
3	10	1
4	9	1
5	8	2
6	8	2
7	7	2
8	9	2
9	11	2
10	8	2
11	11	3
12	11	3
13	8	3
14	8	3
15	9	3
16	10	3
17	14	4
18	11	4
19	10	4
20	9	4
21	11	4
22	12	4
23	12	4
24	10	4

MINITAB: Stat -> ANOVA -> One Way



Output MINITAB

One-Way Analysis of Variance

Analysis of Variance for Sale

Source	DF	SS	MS	F	P
Tr	3	31.21	10.40	3.67	0.029
Error	20	56.62	2.83		
Total	23	87.83			

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	4	11.250	2.630
2	6	8.500	1.378
3	6	9.500	1.378
4	8	11.125	1.553

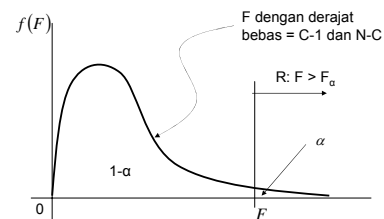
Pooled StDev = 1.683

Dengan metode nilai p:

Nilai $p = 0.029$, sedangkan $\alpha = 0.05$, sehingga nilai $p < \alpha$. Tolak H_0 . Artinya sedikitnya ada satu rata-rata penjualan produk kecantikan yang berbeda dengan yang lainnya

Dengan Metode Nilai Kritis F_α

■ Distribusi F

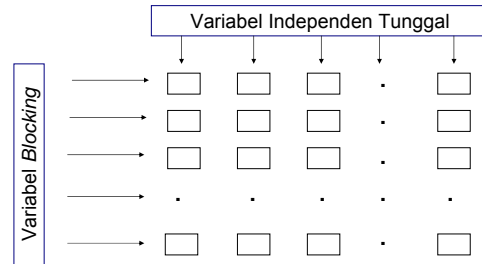


- Pada contoh ini: $F = 3.67$ dan $F_{0.05} = 3.0984$ untuk derajat bebas = C-1 dan N-C. Karena $F > F_{0.05}$, maka tolak H_0 (sama dengan kesimpulan di atas)

Anova Dua Arah (*Two Way Anova*)

- Membandingkan C (>2) populasi sekaligus membandingkan efek blok (*randomized block design*)
- Asumsi:
 - Populasi terdistribusi normal
 - Sampel diambil secara acak dari masing-masing populasi
 - Varians semua populasi sama
- $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_C$
 H_1 : sedikitnya ada 1 rata-rata treatment yang berbeda dengan yang lain
- $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_R$
 H_1 : sedikitnya ada 1 rata-rata blok yang berbeda dengan yang lain

Anova Dua Arah (lanjutan)



Catatan: Setiap sel hanya berisi satu pengamatan

Anova Dua Arah (lanjutan)

Source	DF	SS	MS	F
Block (R =Row)	R - 1	SSR	$MSR = \frac{SSR}{R - 1}$	$F = \frac{MSR}{MSE}$
Treatment (C =Column)	C - 1	SSC	$MSC = \frac{SSC}{C - 1}$	$F = \frac{MSC}{MSE}$
Error	(C-1)(R-1)	SSE	$MSE = \frac{SSE}{(C-1)(R-1)}$	
Jumlah	N - 1	SST		

- N = RC = total banyaknya data yang diamati
- Untuk pengujian efek Blok: derajat bebas F adalah R-1 (pembilang) dan (C-1)(R-1) (penyebut)
- Untuk pengujian efek Treatment: derajat bebas F adalah C-1 (pembilang) dan (C-1)(R-1) (penyebut)

Contoh Aplikasi Anova Dua Arah

- Untuk mengetahui apakah ada pengaruh kemasan (warna dan ukuran kemasan) suatu produk kecantikan terhadap penjualannya, sebuah pabrik alat-alat kecantikan melakukan pengujian dengan membuat kemasan berwarna: merah, kuning, biru, dan hijau dengan ukuran kemasan kecil, sedang, dan besar. Banyaknya produk kecantikan yang terjual selama satu minggu untuk masing-masing kemasan dicatat (terlampir). Dengan menggunakan $\alpha = 5\%$, kesimpulan apakah yang dapat ditarik mengenai pengaruh ukuran kemasan? Kesimpulan apa pula yang dapat ditarik mengenai pengaruh warna kemasan?

Data

	Merah	Kuning	Biru	Hijau	Row	NSale	Ukuran	Warna
Kecil	6	5	6	7	1	6	1	1
					2	7	2	1
					3	9	3	1
					4	5	1	2
					5	9	2	2
Sedang	7	9	6	8	6	8	3	2
					7	6	1	3
					8	6	2	3
					9	10	3	3
					10	7	1	4
Besar	9	8	10	12	11	8	2	4
					12	12	3	4

MINITAB: Stat -> ANOVA -> Two Way

Output MINITAB

Two-way Analysis of Variance

Analysis of Variance for NSale

Source	DF	SS	MS	F	P
Ukuran	2	28.50	14.25	9.00	0.016
Warna	3	6.25	2.08	1.32	0.353
Error	6	9.50	1.58		
Total	11	44.25			

- Efek Blok (ukuran kemasan): $F = 14.25/1.58 = 9.00$, $F_{0.05} = 5.1433$ untuk $df = 2$ dan 6 . Jadi $F > F_{0.05}$, kesimpulan: Tolak H_0 . Artinya: ada pengaruh ukuran terhadap penjualan.
- Efek Treatment (warna kemasan): $F = 2.08/1.58 = 1.32$, $F_{0.05} = 4.7571$ untuk $df = 3$ dan 6 . Jadi $F < F_{0.05}$, kesimpulan: Pertahankan H_0 . Artinya: tidak ada pengaruh warna kemasan terhadap penjualan.
- Metode nilai p juga akan menghasilkan kesimpulan yang sama.

Topik-topik Lanjut

- Regresi Linear Sederhana
- Regresi Berganda
- Deret Waktu
- Statistika Nonparametrik
- dan lain-lain

Daftar Pustaka

- Black, K. 2003. *Business Statistics for Contemporary Decision Making*. 4th Ed. West Publishing Co.
- MINITAB, Inc. 2003. Meet MINITAB Release 14 for Windows
- Lind, D.A. 2002. *Basic Statistics for Business and Economics*. 4nd Ed. McGraw-Hill Companies

Terima kasih

