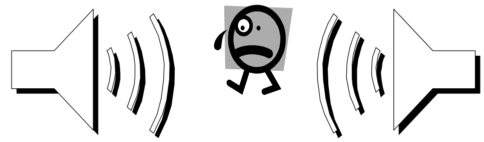
Analisa Kinerja Sistem

Statistik untuk Evaluasi Kinerja

(Chapters 12-15)

Mengapa kita membutuhkan statistik?

1. Noise, noise, noise, noise, noise!

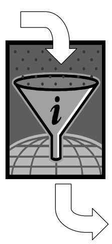


OK – bukan noise seperti yg itu

Mengapa kita membutuhkan statistik?

2. Agregasi data kedalam informasi yang penuh arti

445 446 397 226
388 3445 188 1002
47762 432 54 12
98 345 2245 8839
77492 472 565 999
1 34 882 545 4022
827 572 597 364



$$\bar{x} = \ldots$$

Mengapa kita membutuhkan statistik?

"Impossible things usually don't happen."
- Sam Treiman, Princeton University

- *Statistik* membantu kita untuk meng-kuantifikasi "biasanya."

Apa statistik itu?

- "Kuantitas yang dikomputasi dari sample [data]."

→ Angka tunggal digunakan untuk meringkas koleksi nilai yang lebih besar.

Apa statistik itu?

- "Lies, damn lies, and statistics!"
- "Koleksi dari data kuantitatif."
- "Cabang matematika yg berhubungan dengan koleksi, analisa, interpretasi, dan presentasi sejumlah besar data numerikal."

→ Kita paling tertarik dalam analisa dan interpretation.

Objektif

- Menyediakan latar belakang konseptual yg intuitif untuk beberapa tool standard.
  - Menggambarkan konklusi berarti meskipun ada kemunculan pengukuran noise
  - Memperbolehkan secara benar dan pintar untuk menerapkan teknik dalam situasi yang baru

→ Jangan hanya memasukkan dan menggunakannya dari formula!

Outline

- Introduksi (sudah)
- Dasar-Dasar (berikut)
- Indeks Tendensi Sentral
- Indeks Dispersi
- Membandingkan Sistem
- Misc
- Regresi
- ANOVA

Dasar-Dasar (1/3)

- Event Independen:
  - Satu event tidak mempengaruhi event lain
  - Mengetahui probabilitas satu event tidak mengubah estimasi terhadap yg lainnya
- Fungsi Distribusi Komulatif (atau Kerapatan):
  - $F_x(a) = P(x<=a)$
- Mean (atau Nilai yg Diharapkan):
  - Mean $\mu = E(x) = \Sigma(p_i x_i)$ untuk i atas n
- Varian:
  - Kuadrat dari selisih antara x dan mean
    - $(x- \mu)^2$
  - $Var(x) = E[(x- \mu)^2] = \Sigma p_i (x_i- \mu)^2$
  - Varian ditulis σ. Square root dari varian, σ^2, adalah standard deviasi

Dasar-Dasar(2/3)

- Koefisien Variasi:
  - Rasio deviasi standard terhadap mean
  - C.O.V. = σ / μ
- Kovarian:
  - Derajat dua variabel acak bervariasi satu sama lainnya
  - Cov = $\sigma^2_{xy} = E[(x- \mu_x)(y- \mu_y)]$
  - Dua variabel independen memiliki Cov = 0
- Korelasi:
  - Cov Normalisasi (between –1 and 1)
  - $\rho_{xy} = \sigma^2_{xy} / \sigma_x \sigma_y$
  - Merepresentasikan derajat relationship linier

Dasar-Dasar (3/3)

- Quantile:
  - Nilai x dari CDF pada α
  - Dinyatakan dalam x_α, maka $F(x_\alpha) = \alpha$
  - Terkadang diinginkan .25, .50, .75
- Median:
  - 50-percentile (atau, .5-quantile)
- Mode:
  - Kecenderungan nilai x_i
- Distribusi Normal
  - Distribusi yang paling sering digunakan, kurva "bell"

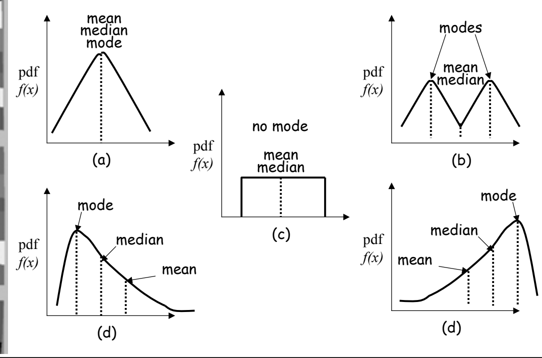
Outline

- Introduksi (sudah)
- Dasar-Dasar (sudah)
- Indeks Tendensi Sentral (berikut)
- Indeks Dispersi
- Membandingkan Sistem
- Misc
- Regresi
- ANOVA

Ringkasan Data oleh Angka Tunggal

- *Indeks tendensi sentral*
- Tiga populer: mean, median, mode
- Mean – sum seluruh observasi, dibagi num
- Median – urut dlm order naik, ambil tengahnya
- Mode – plot histogram dan ambil yg terbesar
- Mean dapat dipengaruhi oleh outliers, sedangkan median atau mode mengabaikan banyak info
- Mean memiliki properti tambahan (mean dari sum adalah sum dari mean), tetapi tidak untuk median atau mode

Relationship Antara Mean, Median, Mode



Petunjuk dalam Memilih Index Tendensi Sentral

- Apakah ini dapat dikategorikan?
  - → ya, gunakan mode
    - Mis: frekuensi microprocessor terbanyak
- Apakah total dipertimbangkan?
  - → ya, gunakan mean
    - Mis: total waktu CPU untuk query (ya)
    - Mis: jumlah window dilayar untuk (tidak)
- Apakah distribusi miring?
  - → ya, gunakan median
  - → tidak, gunakan mean

Contoh untuk Indeks Seleksi Tendensi Sentral

- Resource terbanyak yg digunakan dlm sistem?
  - Kategorikal, maka gunakan mode
- Waktu repons?
  - Total dipertimbangkan, maka gunakan mean
- Load pada komputer?
  - Kemungkinan kemiringan tinggi, maka gunakan median
- Konfigurasi average dari sejumlah disk, jumlah memori, kecepatan network?
  - Kemungkinan miring, maka gunakan median

Salah Penggunaan Mean (1/2)

- Menggunakan mean terhadap nilai yg berbeda secara signifikan
  - Hanya karena mean tepat, tidak dapat dikatakan berguna
    - Mis: dua sample waktu repons, 10 ms dan 1000 ms. Mean adalah 505 ms tetapi tidak berguna.
- Menggunakan mean tanpa memperhatikan kemiringan
  - Tidak merepresentasikan data jika miring
    - Mis: sis. A: 10, 9, 11, 10, 10 (mean 10, mode 10)
    - Mis: sis B: 5, 5, 5, 4, 31 (mean 10, mode 5)

Salah Penggunaan Mean (2/2)

- Multiplying means
  - Mean produk sama dengan produk mean jika dua variabel independen. Tetapi:
    - jika x,y berkorelasi E(xy) != E(x)E(y)
  - Mis: mean user sistem 23, mean proces per user adalah 2. Apakah mean proses sistem? Bukan 46!
  - → Proses ditentukan oleh load, jadi ketika load tinggi maka user lebih sedikit. Maka, harus mengukur proses total dan average
- Mean rasio dengan basis yang berbeda (ditunda)

Mean Geometrik (1/2)

- Mean sebelumnya adalah *mean aritmatika*
  - Digunakan ketika sum dari samples bermakna
  - *Mean Geometrik* ketika produk yg bermakna
- Kalikan nilai n $\{x_1, x_2, \ldots, x_n\}$ dan ambil pangkat $1/n$

  $x = (\Pi x_i)^{1/n}$
- Contoh: mengukur waktu dari peningkatan layer network, dimana 2x layer 1 dan 2x layer 2 sama dengan 4x peningkatan.
- Layer 7 meningkat 18%, 6 13%, 5, 11%, 4 8%, 3 10%, 2 28%, 1 5%
- Maka, mean geometrik per layer:
  - $[(1.18)(1.13)(1.11)(1.08)(1.10)(1.28)(1.05)]^{1/7} - 1$
  - Peningkatan average per layer adalah 0.13, or 13%

Mean Geometrik (2/2)

- Contoh lain metrik yang berlaku dalam pendekatan multiplikatif:
  - Cache hit ratio pada beberapa level
    - Dan cache miss ratio
  - Persentasi peningkatan kinerja antara versi berikutnya
  - Average error rate per hop pada path multi-hop dalam network

Mean Harmonik (1/2)

- Mean harmonik samples $\{x_1, x_2, \ldots, x_n\}$ adalah:

  $n / (1/x_1 + 1/x_2 + \ldots + 1/x_n)$
- Digunakan ketika mean aritmatika berlaku untuk 1/x
- Mis: pengukuran elapsed processor benchmark dari m instruksi. Instruksi ke i memakan t_i detik. MIPS x_i adalah m/t_i
  - Karena sum dari instruksi bermakna, dapat menggunakan mean harmonik
  - $= n / [1/(m/t_1) + 1/(m/t_2) + \ldots + 1/(m/t_n)]$
  - $= m / [(1/n)(t_1 + t_2 + \ldots + t_n)$

Mean Harmonik (2/2)

- Mis: if different benchmarks (m_i), then sum of m_i/t_i does not make sense
- Instead, use weighted harmonic mean

  $n / (w_1/x_1 + w_2/x_2 + \ldots + w_3/x_n)$
  - where $w_1 + w_2 + .. + w_n = 1$
- In example, perhaps choose weights proportional to size of benchmarks
  - $w_i = m_i / (m_1 + m_2 + .. + m_n)$
- So, weighted harmonic mean

  $(m_1 + m_2 + .. + m_n) / (t_1 + t_2 + .. + t_n)$
  - Reasonable, since top is total size and bottom is total time

Mean Rasio (1/2)

- Set dejumlah n rasio, bagaimana meringkasnya?
- Ini, jika sum numerator dan sum denominator keduanya memiliki makna, rasio average adalah rasion average

  Average$(a_1/b_1, a_2/b_2, \ldots, a_n/b_n)$

  $= (a_1 + a_2 + \ldots + a_n) / (b_1 + b_2 + \ldots + b_n)$

  $= [(\Sigma a_i)/n] / [(\Sigma b_i)/n]$
- Umum digunakan dalam menghitung mean utilisasi resource (contoh berikutnya)

Mean Rasio (2/2)

- Utilisasi CPU:
  - Untuk durasi 1 sibuk 45%, 1 %45, 1 45%, 1 45%, 100 20%
  - Sum 200%, mean != 200/5 or 40%
    - Denominator basis (durasi) tidak dapat dibandingkan
  - mean = sum kesibukan CPU/ sum durasi
  - = (.45+.45+.45+.45+20) / (1+1+1+1+100)
  - = 21%

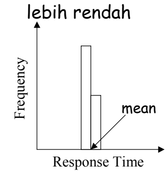
Outline

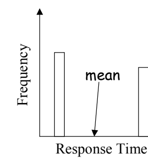
- Introduksi (sudah)
- Dasar-Dasar (sudah)
- Indices of Central Tendency (sudah)
- Indeks dispersi (berikut)
- Membandingkan Sistem
- Misc
- Regresi
- ANOVA

Meringkas Variabilitas (1/2)

> *"Then there is the man who drowned crossing a stream with an average depth of six inches."* – W.I.E. Gates

- Meringkas dengan menggunakan angka tunggal cukup jarang dilakukan → perlu pernyataan tentang *variabilitas*
  - Jika dua sistem meiliki mean yang sama, cenderung memilih salah satu yang memiliki variabilitas yang lebih rendah





Meringkas Variabilitas (2/2)

- *Indeks Dispersi*
  - *Range* – nilai min dan max values terobservasi
  - *Variance* atau *deviasi stantard*
  - 10- dan 90-*percentiles*
  - Range inter*interquartile (semi)*
  - *Mean deviasi absoluteabsolute deviation*

(Masing-masing akan dibicarakan berikut)

Range

- Mudah dilacak
- Rekam max dan min, kurangi
- Kebanyakan, tidak begitu berguna:
  - Minimum mungkin nol
  - Maximum dapat berasal dari outlier
    - Event sistem tidak berhubungan dengan fenomena yg diobservasi
  - Maximum lebih besar dengan banyaknya sample, jadi tidak ada titik "stabil"
- Akan tetapi, jika sistem dibatasi, untuk sample besar, range mungkin memberikan batas

Variance Sample

- *Variance* sample (dapat tanpa "sample" jika maknanya jelas)
  - $s^2 = [1/(n-1)] \Sigma(x_i - \underline{x})^2$
- Ingat (n-1) karena hanya n-1 independen
  - Juga disebut *derajat kebebasan*
- Problem utama adalah pada unit kuadrat, merubah unit merubah kuadrat jawabnya
  - Mis: waktu respons .5, .4, .6 seconds

  Variance = 0.01 seconds squared atau 10000 msecs squared

Deviasi Standard

- Maka, gunakan *deviasi standard*
  - s = sqrt(s^2)
  - Unit sama dengan *mean*, maka dapat dibandingkan dengan *mean*
- Mis: waktu repons .5, .4, .6 seconds
  - stddev .1 seconds or 100 msecs
  - Dapat dibandingkan masing-masing dengan mean
- Rasio *deviasi standard* dengan *mean*?
  - Disebut *Coefficient of Variation* (C.O.V.)
  - Hilangkan unitnya dan tunjukkan besarnya
  - Mis: diatas adalah 1/5th (or .2) untuk masing-masing unit

Percentiles/Quantile

- Mirip dengan range
- Nilai dalam bentuk persen (atau fraksi)
  - 90-percentile, 0.9-quantile
  - Untuk α–quantile, urut dan ambil ke [(n-1)α+1]
    - [] artinya dibulatkan ke integer terdekat
- 25%, 50%, 75% → *quartiles* (Q1, Q2, Q3)
  - Catatan, Q2 juga median
- Range Q3 – Q1 adalah range *interquartile*
  - ½ dari (Q3 – Q1) adalah range *semi-interquartile*

Deviasi Absolut Mean

- $(1/n) \Sigma|x_i - \underline{x}|$
- Mirip dengan standar deviasi, tetapi tidak membutuhkan multiplikasi atau akar kuadrat
  - (Outliers tidak dikuadratkan)
- Maka, bagaimanan So, seberapa rawannya indeks dispersi dari outlier?

Ringkasan Indeks Dispersi

- Ranking pengaruh dari outliers
  - Range beresiko
  - Variance (deviasi standar)
  - Deviasi absolut mean
  - Range semi-interquartile resistant
- Gunakan semi-interquantile (SIQR) untuk indeks dispersi kapanpun menggunakan median sebagai indeks tendensi sentral
- Catatan, semua hanya dapat diterapkan untuk data kuantitatif
  - Untuk kualitatif (kategorikal) berikan banyaknya kategori untuk percentile sample yang ada

Contoh Indeks Dispersi

(Sorted)
CPU Time

1.9	3.9
2.7	3.9
2.8	4.1
2.8	4.1
2.8	4.2
2.9	4.2
3.1	4.4
3.1	4.5
3.2	4.5
3.2	4.8
3.3	4.9
3.4	5.1
3.6	5.1
3.7	5.3
3.8	5.6
3.9	5.9

- Pertama, urutkan
- Median = [1 + 31*.5] = 16^{th} = 3.2
- Q1 = 1 + .31 * .25 = 9^{th} = 3.9
- Q3 = 1 + .31*.75 = 24^{th} = 4.5
- *SIQR* = (Q3–Q1)/2 = .65
- *Variance* = 0.898
- *Stddev* = 0.948
- *Range* = 5.9 – 1.9 = 4

Memilih Indeks Dispersi

- Apakah distribusi dibatasi
  - Ya? → gunakan range
- Tidak? Apakah distribusi simetrik unimodal unimodal symmetric?
  - Ya? → Gunakan C.O.V.
- Tidak?
  - Gunakan percentiles atau SIQR
- Buka aturan yang cepat dan tepat, tetapi hanya *panduan*
  - Mis: dispersi load network. Dapat menggunakan range bahkan C.O.V. Tetapi ingin mengakomodasi 90% atau 95% load, maka gunakan percentile. Mirip power supply.

Menentukan Distribusi Data

- Informasi ringkasan tambahan mungkin adalah distribusi data
  - Mis: Mean I/O disk 13, variance 48. Ok. Mungkin akan lebih berguna untuk dengan menyatakan bahwa data *terdistribusi secara uniform* antara 1 sampai 25.
  - Ditambah, distribusi berguna untuk simulasi mendatang atau pemodelan analitik
- Bagaimana menyatakan distribusi?
  - Pertama, plot histogram

Histograms

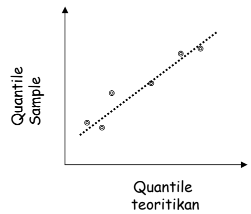
- Perlu: max, min, ukuran buckets
- Menentukan ukuran sel merupakan masalah
  - Terlalu sedikit, sulit melihat distro
  - Terlalu banyak, kehilangan distro
  - Guideline:
    - Jika sel > 5 maka pisahkan

Cell	#	Histogram (size 1)
1	1	X
2	5	XXXXX
3	12	XXXXXXXXXXXX
4	9	XXXXXXXXX
5	5	XXXXX

Cell	#	Histogram (size .2)
1.8	1	X
2.6	1	X
2.8	4	XXXX
3.0	2	XX
3.2	3	XXX
3.4	1	X
3.6	2	XX
3.8	4	XXXX
4.0	2	XX
4.2	2	XX
4.4	3	XXX
4.8	2	XX
5.0	2	XX
5.2	1	X
5.6	1	X
5.8	1	X

Distribusi Data

- Alternatif, plot quantile terobservasi vs quantile teoritikal
  - y_i observasi, x_i teoritikal
  - Jika distribusi fit, akan mendapatkan garis



Perlu inversi CDF: $q_i = F(x_i)$, atau $x_i = F^{-1}(q_i)$

Dimana F^{-1}? Table 28.1 Untuk beberapa distribusi

Distribusi Normal
$x_i = 4.91[q_i^{0.14} - (1-q_i)^{0.14}]$

Table 28.1

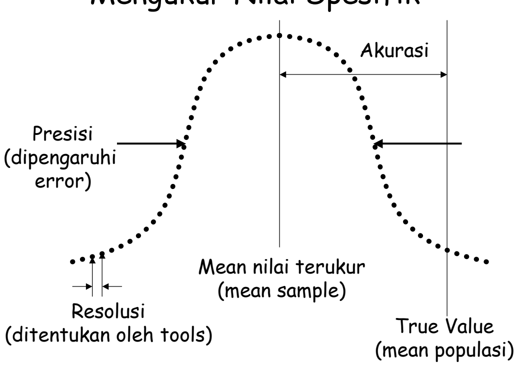
Distribution	CDF $F(x)$	Inverse
Exponential	$1 - e^{-x/a}$	$-a\ln(u)$
Extreme value	$1 - e^{-e^{\frac{x-a}{b}}}$	$a + b\ln\ln u$
Geometric	$1 - (1-p)^x$	$\left\lceil\frac{\ln(u)}{\ln(1-p)}\right\rceil$
Logistic	$1 - \frac{1}{1+e^{\frac{x-\mu}{b}}}$	$\mu - b\ln(\frac{1}{u} - 1)$
Pareto	$1 - x^{-a}$	$1/u^{1/a}$
Weibull	$1 - e^{(x/a)^b}$	$a(\ln u)^{1/b}$

Distribusi normal:
$x_i = 4.91[q_i^{0.14} - (1-q_i)^{0.14}]$

Outline

- Introduksi (sudah)
- Dasar-dasar (sudah)
- Indeks Tendensi Sentral Tendency (sudah)
- Indeks Dispersi (sudah)
- Membandingkan Sistem (berikut)
- Misc
- Regresi
- ANOVA

Mengukur Nilai Spesifik



Membandingkan Sistem Menggunakan Data Sample

"Statistics are like alienists – they will testify for either side." – Fiorello La Guardia

- Kata "sample" berasal dari akar yang sama dengan "example"
- Juga,satu sample tidak membuktikan teori, akan tetapi lebih merupakan example
- Pada dasarnya, pernyataan pasti tidak dapat dibuat tentang karakteristik dari semua sistem
- Melainkan, membuat pernyataan probabilistik tentang range dari kebanyakan sistem
  - *Confidence intervals*

Sample versus Populasi

- Say kita men-generate 1-juta angka random
  - mean μ dan stddev σ.
  - μ adalah *mean populasi*
- Letakkan dalam deretan sample *n*
  - Sample $\{x_1, x_2, \ldots, x_n\}$ memiliki mean $\bar{x}$, stddev s
- $\bar{x}$ cenderung berbeda dari μ!
  - Dengan banyak sample, $\bar{x}_1 \mathrel{!=} \bar{x}_2 \mathrel{!=} \ldots$
- Secara tipikal, μ tidak diketahui dan mungkin sulit untuk diketahui
  - Melainkan, dapatkan estimasi μ dari $\bar{x}_1, \bar{x}_2, \ldots$

Confidence Interval untuk Mean

- Cari probabilitas μ dalam interval $[c_1, c_2]$
  - $\text{Prob}\{c_1 \le \mu \le c_2\} = 1-\alpha$
    - (c1, c2) adalah *confidence interval*
    - α adalah *level signifikansi*
    - $100(1-\alpha)$ adalah *confidence level*
- Biasanya ingin α kecil maka confidence level 90%, 95% atau 99% (lebih banyak lagi nanti)
- Katakan, α =0.1. dapat mengambil *k* sample, temukan mean sample, sort urutkan
  - Interval: $[1+0.05(k-1)]^{th}$ dan $[1+0.95(k-1)]^{th}$
    - 90% confidence interval
- Kita harus mengambil *k* sample, masing-masing yg berukuran *n*?

Teorima Limit Sentral

Sum of a "large" number of values from any distribution will be normally distributed.

- Tidak memerlukan banyak sample. Satu sudah cukup.

$$\bar{x} \sim N(\mu, \sigma/\text{sqrt}(n))$$

- Standard error = $\sigma/\text{sqrt}(n)$
  - Sejalan dengan kenaikan ukuran sample *n*, error turun
- Maka, $100(1-\alpha)$% confidence interval untuk mean populasi adalah:

$$(\bar{x}-z_{1-\alpha/2}s/\text{sqrt}(n),\ \bar{x}+z_{1-\alpha/2}s/\text{sqrt}(n))$$

- Dimana $z_{1-\alpha/2}$ adalah $(1-\alpha/2)$-quantile dari normal unit (Table A.2 ada dalamin appendix, A.3 umum)

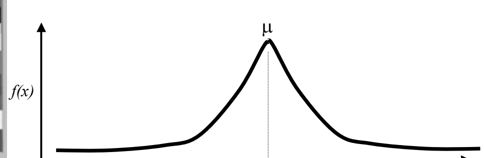
Contoh Confidence Interval

(Sorted)
CPU Time

1.9	3.9
2.7	3.9
2.8	4.1
2.8	4.1
2.8	4.2
2.9	4.2
3.1	4.4
3.1	4.5
3.2	4.5
3.2	4.8
3.3	4.9
3.4	5.1
3.6	5.1
3.7	5.3
3.8	5.6
3.9	5.9

- $\bar{x}$ = 3.90, stddev s=0.95, *n*=32
- 90% confidence interval untukl mean populasi (μ):
  3.90 +- (1.645)(0.95)/sqrt(32)
  = (3.62, 4.17)
- Dengan 90% confidence, μ dalam interval tsb. Kemungkinan error 10%.
  - Jika mengambil sample 100 dan membuat confidence intervals seperti diatas, dalam 90 sus interval termasuk μ dan dlm 10 kasus tdk termasuk μ

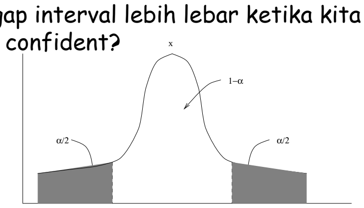
Arti dari Confidence Interval



Sample	Termasuk μ?
1	ya
2	ya
3	tdk
…	
100	ya
Total	ya $\ge 100(1-\alpha)$
Total	tdk $\le 100\alpha$

Bagaimana Interval Berubah?

- 90% CI = [6.5, 9.4]
  - 90% kemungkinan nilai real antara 6.5, 9.4
- 95% CI = [6.1, 9.7]
  - 95% kemungkinan nilai real antara 6.1, 9.7
- Mengap interval lebih lebar ketika kita lebih confident?



Bagaimana jika *n* tidak besar?

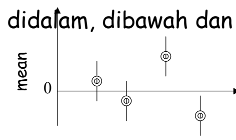
- Datas hanya berlaku untuk sample besar, 30+
- Untuk *n* yg lebih kecil, hanya dapat membuat confidence interval jika observasi berasal dari populasi yg terdistribusi secara normal
  - Apakah ini benar untuk sistem komputer?

  $(\underline{x}-t_{[1-\alpha/2;n-1]}s/sqrt(n), \underline{x}+t_{[1-\alpha/2;n-1]}s/sqrt(n))$
- Table A.4. (Student's t distribution. "Student" adalah nama anonimus)

Again, n-1 degrees freedom

Pengujian untuk *Zero Mean*

- Umum untuk memeriksa jika nilai terukur secara signifikan berbeda dibanding nol
- Dapat menggunakan confidence interval dan kemudian memeriksa apakah 0 dalam interval.
- Mungkin didalam, dibawah dan diatas



Catatan, dapat diperluas dgn menyertakan pengujian setiap nilai *a*

Contoh: Pengujian untuk Zero Mean

- Tujuh workload
- Perbedaan waktu CPU dari dua algoritma

  {1.5, 2.6, -1.8, 1.3,-0.5, 1.7, 2.4}
- Dapatkah dikatakan dengan 99% confidence bahwa satu algoritma lebih unggul dari lainnya?
- $n = 7$, $\alpha = 0.01$
- mean = 7.20/7 = 1.03
- variance = 2.57 maka stddev = sqrt(2.57) = 1.60
- CI = 1.03 +- tx1.60/sqrt(7) = 1.03 +- 0.605t
- 1 - $\alpha/2$ = .995, maka $t[0.995;6]$ = 3.707 (Table A.4)
- 99% confidence interval = (-1.21, 3.27)

→ Dengan 99% confidence, kinerja algoritma adalah identik

Membandingkan Dua Alternatif

- Sering ingin membandingkan sistem
  - Sistem A dengan sistem B
  - Sistem "sebelum" dan sistem "sesudah"
- Observasi berpasangan
- Observasi tidak berpasangan
- Uji pendekaytan visual

Observasi Berpasangan

- jika *n* eksperimen dimana korespondensi 1-ke-1 dari uji pada A dengan uji pada B maka *berpasangan*
  - (Jika tidak ada korespondensi, maka *tidak berpasangan)*
- Memperlakukan dua sample sebagai satu sample pasangan *n*
- Untuk tiap pasangan, hitung perbedaannya
- Buat confidence interval untuk perbedaannya
- Jika CI termasuk 0, maka sistem tidak berbeda secara signifikan

Contoh: Observasi Berpasangan

- Mengukur ukuran workload berbeda pada A and B

  {(5.4, 19.1), (16.6, 3.5), (0.6,3.4), (1.4,2.5), (0.6, 3.6) (7.3, 1.7)}
- Apakah satu sistem lebih baik dari lainnya?
- Enam perbedaan terobservasi
  - {-13.7, 13.1, -2.8, -1.1, -3.0, 5.6}
- Mean = -.32, stddev = 9.03
- CI = -0.32 +- t[sqrt(81.62/6)] = -0.32 +- t(3.69)
- .95 quantile dari t dengan 5 derajat kebebasan = 2.015
- 90% confidence interval = (-7.75, 7.11)
- Oleh karena itu, dua sistem tidak berbeda

Observasi Tidak Berpasangan

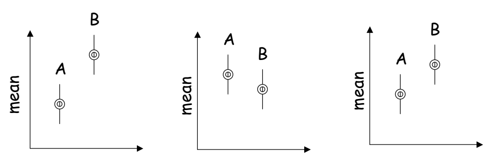
- Sistem A, B dengan sample n_a and n_b
- Hitung mean sample: $\underline{x}_a$, $\underline{x}_b$
- Hitung devciasi standard: s_a, s_b
- Hitung perbedaan mean: $\underline{x}_a$-$\underline{x}_b$
- Hitung stddev dari perbedaan mean:
  - S = sqrt($s_a^2/n_a + s_b^2/n_b$)
- Hitung derajat kebebasan efektif
- Hitung confidence interval
- Jika interval termasuk 0, maka tidak ada perbedaan yg signifikan

Contoh: Observasi Tidak Berpasangan

- Waktu prosesor untuk task pada dua sistem
  - A: {5.36, 16.57, 0.62, 1.41, 0.64, 7.26}
  - B: {19.12, 3.52, 3.38, 2.50, 3.60, 1.74}
- Apakah kedua sistem secara signifikan berbeda?
- Mean x_a = 5.31, s_a^2 = 37.92, n_a=6
- Mean x_b = 5.64, s_b^2 = 44.11, n_b =6
- Perbedaab mean difference x_a-x_b = -0.33
- Stddev dari perbedaan mean = 3.698
- t adalah 1.71
- 90% confidence interval = (-6.92, 6.26)
  - Tidak berbeda

Uji Pendekatan Visual

- Hitung confidence interval untuk mean
- Lihat apakah terjadi overlap



CIs not overlap → A higher than B

CIs do overlap and Mean of one in another → Not different

CIs do overlap but mean of one not in another → Do t test

Example: Uji Pendekatan Visual

- Waktu prosesor untuk task pada dua sistem
  - A: {5.36, 16.57, 0.62, 1.41, 0.64, 7.26}
  - B: {19.12, 3.52, 3.38, 2.50, 3.60, 1.74}
- t-value pada 90%, 5 adalah 2.015
- 90% confidence intervals
  - A = 5.31 +-(2.015)sqrt(37.92/6) = (0.24,10.38)
  - B = 5.64 +-(2.015)sqrt(44.11/6) = (0.18,11.10)
- Dua confidence interval overlap dan mean satunya berada dalam interval lainnya. Oleh karena itu dua sistem tidak berbeda tanpa unpaired t test

Outline

- Introduksi
- Dasar-dasar
- Indeks Tendensi Sentral
- Indeks Dispersi
- Membandingkan Sistem
- Misc
- Regresi
- ANOVA

Confidence Level mana yg Digunakan?

- Sering melihat 90% atau 95% (atau bahkan 99%)
- Pilihan berdasarkan pada loss jika parameter populasi diluar atau gain jika parameter didalam
  - Jika loss lebih tinggi dibandingkan gain, gunakan high confidence
  - Jika loss rendah dibandingkan gain, gunakan low confidence
  - Jika loss diabaikan, low juga baik
- Contoh:
  - Tiket lotre \$1, bayar \$5 million
  - Kemungkinan menang 10^{-7} (1 dalam 10 juta)
  - Untuk memenangkan dgn 90% confidence, memerlukan 9 juta tiket
    - Tidak ada seorangpun akan membeli banyak tiket!
  - Maka, kebanyakan orang puas dengan 0.01% confidence

Uji Hipotesa

- Uji hipotesa biasanya diterima/ditolak
  - Dapat dilakukan dengan menggunakan confidence intervals
- Plus, interval memberikan kita lebih … interval
- Mis: sistem A dan B
  - CI (-100,100) kita dapat katakan "tidak ada perbedaan"
  - CI(-1, 1) dikatakan lebih "tidak ada perbedaan
- Confidence intervals lebih mudah menjelaskan karena unit sama dengan apa yang diukur
  - Mis: lebih berguna untuk mengetahui range 100 sampai 200 dibandingkan probabilitas nilai dibawah 110 adalah 3%

Confidence Intervals Satu-sisi

- Pada 90% confidence, 5% kemungkina lebih rendah dari limit dan 5 % lebih tinggi dari limit
- Terkadang, hanya ingin melakukan perbandingan satu sisi
  - Katakan, uji jika mean lebih besar dri nilai

    $(\underline{x}-t_{[1-\alpha;n-1]}s/sqrt(n),\underline{x})$
  - Gunakan $1-\alpha$ selain $1-\alpha/2$
- Begitu juga (tetapi dgn +) untuk batas atas confidence
- Dapt menggunakan nilai-z jika lebih dari 30

Confidence Intervals untuk Proporsi

- Variabel kategorikal terkadang memiliki probabilitas dengan tiap kategori → disebut *proporsi*
  - Ingin CI atas proporsi
- Tiap sample n observasi memberikan proporsi sample (katakan, dari tipe 1)
  - n_1 dari n observasi adalah tipe 1

    $p = n_1 / n$
- CI untuk p: $p+-z_{1-\alpha/2}sqrt(p(1-p)/n)$
- Hanya valid jika $np \geq 10$
  - Selain itu, terlalu komplikasi. Lihat buku statistik.

Contoh: CI untuk Proporsi

- 10 dari 1000 halaman tercetak tidak terbaca

  $p = 10/1000 = 0.01$
- Karena $np \geq 10$ maka dapat menggunakan persamaan sebelumnya

CI = p +- z(sqrt(p(1-p)/n))
= 0.01 +- z(sqrt(0.01(0.99)/1000)
= 0.01 +- 0.003z

90% CI = 0.01 +- (0.003)(1.645) = (0.005, 0.015)

- Maka, pada 90% confidence dapat dikatakan bahwa 0.5% - 1.5% halaman tidak terbaca.
  - Terdapat 10% kemungkinan bahwa pernyataan ini error

Menentukan Ukuran Sample

- Semakin tinggi ukuran sample, semakin besar confidence dalam kesimpulan
  - CI lebih ketat karena dibagi dengan sqrt(n)
  - Tetapi semakin banyak sample semakin menuntut resource (waktu)
- Goalnya adalah menemukan ukuran sample terkecil untuk menghasilkan confidence yg diinginkan dalam hasilnya
- Metode:
  - Sekelompok kecil pengukuran
  - Gunakan untuk mengestimasi variance
  - Gunakan untuk menentukan ukuran sample untuk keakurasian

Ukuran Sample untuk Mean

- Jika menginginkan kinerja mean dengan akurasi +-r% at $100(1-\alpha)$% confidence
- Ukuran sample diketahui n, CI adalah

  $\underline{x} +- z(s/sqrt(n))$
- Maka CI adalah $[\underline{x}(1-r/100), \underline{x}(1+r/100)]$

  $\underline{x} +- z(s/sqrt(n)) = \underline{x}(1 +- r/100)$

  $z(s/sqrt(n)) = \underline{x}(r/100)$

  $n = [(100zs)/(r\underline{x})]^2$

Contoh: Ukuran Sample untuk Mean

- Uji awal:
  - Waktu respons 20 detik
  - stddev = 5 detik
- Berapa banyak pengulangan untuk mendapatkan akurasi waktu dalam 1 detik pada 95% confidence

$\underline{x}$=20, s=5, z=1.960, r=5 (1 sec is 5% of 20)

$$n = [(100 \times 1.960 \times 5) / (5 \times 20)]^2$$
$$= (9.8)^2$$
$$= 96.04$$

- Maka, total 97 observasi dibutuhkan
- Dapat diperluas untuk proporsi (tidak ditunjukkan)

Contoh: Ukuran Sample untuk Membandingkan Alternatif

- Membutuhkan non-overlapping confidence intervals
- Algoritma A hilang 0.5% paket dan B hilang 0.6%
- Berapa jumlah paket yang dibutuhkan untuk menyatakan bahwa alg A lebih baik dr alg B pada 95%?

CI untuk A: $0.005 \pm 1.960[0.005(1-0.005)/n]^{\frac{1}{2}}$

CI untuk B: $0.006 \pm 1.960[0.006(1-0.006)/n]^{\frac{1}{2}}$

- Perlu tepi atas A tidak overlap dgn tepi bawah B

$$0.005 + 1.960[0.005(1-0.005)/n]^{\frac{1}{2}} <$$
$$0.006 - 1.960[0.006(1-0.006)/n]^{\frac{1}{2}}$$

Solusi untuk n: $n > 84{,}340$

- Maka, dibutuhkan 85000 paket

Ringkasan

- Statistics adalah tools
  - Membantu menggambarkan konklusi
  - Meringkas dgn cara yang berarti dengan kehadiran noise
- Indeks tendensi sentral dan Indeks dispersi sentral
  - Meringkas data dengan sedikit angka
- Confidence intervals

Outline

• Introduksi	(sudah)
• Dasar-dasar	(sudah)
• Indeks Tendensi Sentral	(sudah)
• Indeks Dispersi	(sudah)
• Membandingkan Sistem	(sudah)
• Misc	(sudah)
• Regresi	(berikut)
• ANOVA	

Regresi

> *"I see your point … and raise you a line."*
> – Elliot Smorodinksy

- Mahal (dan terkadang impossible) untuk mengukur kinerja dengan segala kemungkinan nilai input
- Selain itu, mengukur kinerja untuk input terbatas dan menggunakan untuk menghasilkan model dengan nilai input jangkauan tertentu
  - Bangun model *regresi*

Regresi Linier (1/2)

- Menangkap hubungan linier antara nilai input dan respons
  - *Minimisasi least-squares*
- Dengan bentuk:

$$y = a + bx$$

- Dimanae x input, y respons dan kita ingin mengetahui
- Jika y_i diukur untuk input input x_i, kemudian tiap pasangan (x_i, y_i) dapat ditulis:

$$y_i = a + bx_i + e_i$$

- dimana e_i adalah residual (error) untuk model regresi

Regrasi Linier (2/2)

- The sum of the errors squared:

$$SSE = \Sigma e_i^2 = \Sigma(y_i - a - bx_i)^2$$

- Cari a dan b yang meminimalkan SSE
- Ambil derivatif berdasarkan a dan kemudian b dan kemudian diset nol

$$na + b\Sigma x_i = \Sigma y_i \qquad (1)$$

$$a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i$$

(two equations in two unknowns)

- Jawab untuk b:

$$b = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

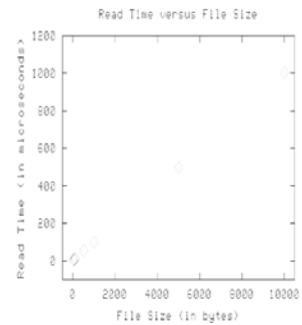
- Menggunakan (1) dan menyelesaikan a:

$$a = \underline{y} - b\underline{x}$$

Contoh Regresi Linier (1/3)

File Size (bytes)	Time (μsec)
10	3.8
50	8.1
100	11.9
500	55.6
1000	99.6
5000	500.2
10000	1006.1

Bangun model regresi linier waktu untuk membaca file dengan ukuran byte



Contoh Regresi Linier (2/3)

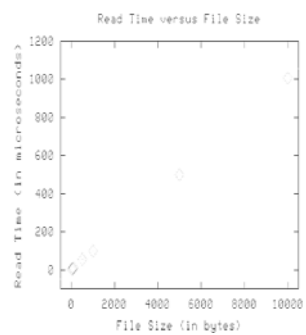
File Size (bytes)	Time (μsec)
10	3.8
50	8.1
100	11.9
500	55.6
1000	99.6
5000	500.2
10000	1006.1

Bangun model regresi linier waktu untuk membaca file dengan ukuran byte

- $\Sigma x_i = 16,660.0$
- $\Sigma y_i = 1685.3$
- $\Sigma x_i y_i = 12,691,033.0$
- $\Sigma x_i^2 = 126,262,600.0$
- $\underline{x} = 2380$
- $\underline{y} = 240.76$
- $b = \dfrac{(7)(12691033) - (16660)(1685.3)}{(7)(126262600) - (16660)^2}$
- $a = 240.76 - .1002(2380)$
  $= 2.24$
- $y = 2.24 + 0.1002x$

Contoh Regresi Linier (2/3)

File Size (bytes)	Time (μsec)
10	3.8
50	8.1
100	11.9
500	55.6
1000	99.6
5000	500.2
10000	1006.1



$y = 2.24 + 0.1002x$

Mis: prediksi waktu untuk membaca 3 k file 303 μdetik

Confidence Intervals untuk Parameter Regresi (1/3)

- Karena parameter a dan b berbasis pada nilai pengukuran dengan error, nilai terprediksi (y) juga memiliki kemungkinan error
- Dapat menderivasi confidence intervals untuk a dan b
- Pertama, butuh estimasi variance dari a dan b

$$s^2 = SSE / (n-2)$$

  - Dengan n pengukuran dan dua variabel, derajat kebebasan adalah n-2
- Perluasan SSE

$$= \Sigma e_i^2 = \Sigma(y_i - a - bx_i)^2 = \Sigma[(y_i - \underline{y}) - b(x_i - \underline{x})]^2$$

Confidence Intervals untuk Parameter Regresi (2/3)

- Membantu untuk merepresentasikan SSE sebagai:

$$SSE = S_{yy} - 2bS_{xy} + b^2 2S_{xx} = S_{yy} - bS_{xy}$$

- Dimana

$$S_{xx} = \Sigma(x_i - \underline{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2 / n$$

$$S_{yy} = \Sigma(y_i - \underline{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2 / n$$

$$S_{xy} = \Sigma(x_i - \underline{x})(y_i - \underline{y}) = \Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i) / n$$

- So, $s^2 = SSE / (n-2)$

$$= S_{yy} - bS_{xy} / (n-2)$$

Confidence Intervals untuk Parameter Regresi (3/3)

- Conf interval untuk slope (b) dan y perpotongan (a):

$$[b_1,b_2] = b \pm t_{[1-\alpha/2;n-2]} s \,/\, \mathrm{sqrt}(S_{xx})$$

$$[a_1,a_2] = a \pm t_{[1-\alpha/2;n-2]} \frac{s \times \mathrm{sqrt}(\Sigma x_i^2)}{\mathrm{sqrt}(nS_{xx})}$$

- Akhirnya, untuk prediksi y_p dapat menentukan interval $[y_{p1}, y_{p2}]$:

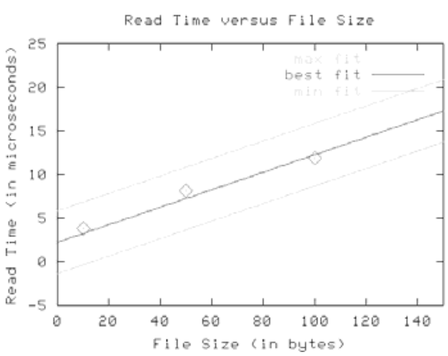
$$= y_p \pm t_{[1-\alpha/2;n-2]} s \times \mathrm{sqrt}\,(1 + 1/n + (x_p-\underline{x})^2/S_{xx})$$

Contoh Conf Interval Regresi (1/2)

$\boxed{y = 2.24 + 0.1002x}$

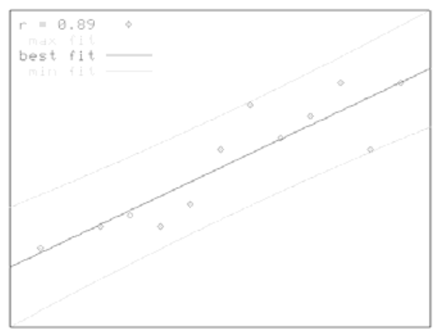
- Σx_i = 16,660.0
- Σy_i = 1685.3
- $\Sigma x_i y_i$ = 12,691,033.0
- Σx_i^2 = 126,262,600.0
- $\underline{x}$ = 2380
- $\underline{y}$ = 240.76
- $b = \frac{(7)(12691033) - (16660)(1685.3)}{(7)(126262600) - (16660)^2}$
- a = 240.76-.1002(2380) = 2.24
- y = 2.24 + 0.1002x
- $S_{xx} = 126262600 - 16660^2/7$ = 86,611,800
- $S_{yy} = 1275670.43 - (1685.3)^2 / 7$ = 869,922.42
- S_{xy} = 12691033-(16660)(1685.3)/7 = 8,680,019
- $s^2 = \frac{869922.42 - 0.1002(8680019)}{(7-2)}$
- Std dev s = sqrt(36.9027) = 6.0748
- 90% conf interval
  - $[b_1,b_2]$ = [0.099, 0.102]
  - $[a_1,a_2]$ = [-3.35, 7.83]

Contoh Conf Interval Regresi (2/2)

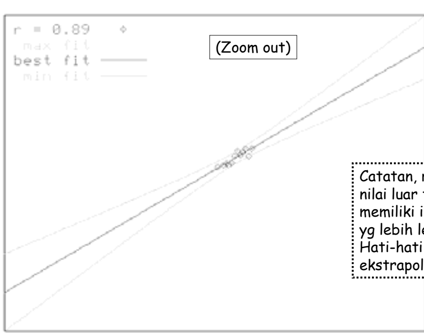


(Zoom)

Contoh Conf Interval Regresi Lainnya (1/2)

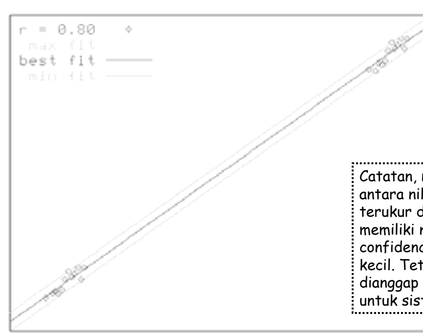


Contoh Conf Interval Regresi Lainnya (2/2)



Catatan, range nilai luar terukur memiliki interval yg lebih lebar!, Hati-hati dengan ekstrapolasi

Another Regression Conf Interval Example



Catatan, nilai antara nilai terukur dapat memiliki nilai confidence kecil. Tetapi dianggap wajar untuk sistem

Korelasi

- Setelah mengembangkan model regresi, berguna mengetahui seberapa baik persamaan regresi cocok untuk data
  - *Koefisien Determinasi*
    - Menyatakan berapa banyak total variasi ygn dijelaskan oleh model linier
  - *Koefisien Korelasi*
    - Square root dari koefisien determinasi

Koefisien Determinasi

- Sebelumnya: SSE = S_{yy} – bS_{xy}
- Jika: SST = S_{yy} dan SSR = bS_{xy}
- Maka: SST = SSR + SSE
  - Total variasi (SST) memiliki dua komponen
    - Porsi SSR dijelaskan oleh regrasi
    - SSE adalah error model (jarak dari garis)
- Fraksi dari total variasi dijelaskan oleh garis model:

  r^2 = SSR / SST = (SST – SSE) / SST
  - Disebut *koefisien determinasi*
- Seberapa "baik" model regrasi ini? Secara kasar:
  - 0.8 <= r^2 <= 1 strong
  - 0.5 <= r^2 < 0.8 medium
  - 0 <= r^2 < 0.5 weak

Koefisien Korelasi

- Square root dari koefisien determinasi adalah koefisiean korelasi. Atau:

  r = S_{xy} / sqrt($S_{xx}S_{yy}$)
- Catatan, secara ekuivalen:

  r = b sqrt(S_{xx}/S_{yy}) = sqrt(SSR/SST)
  - Dimana b = S_{xy}/S_{xx} adalah slope garis model regresi
- Nilai r memiliki range antara –1 and +1
  - +1 adalah relasi positif linier sempurna
    - Perubahan dalam x menghasilkan korespondensinya dalam y
  - -1 adalah relasi negatif linier sempurna

Contoh Korelasi

- Dari Ukuran Read vs. Model waktu, korelasi:

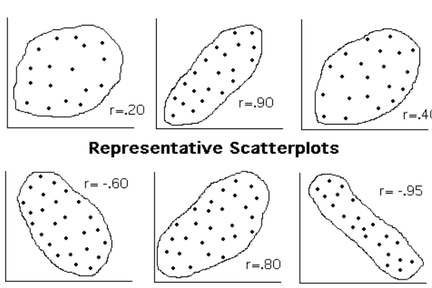
  r = b sqrt(S_{xx}/S_{yy})

  = 0.1002 sqrt(86,611,800 / 869,922.4171)

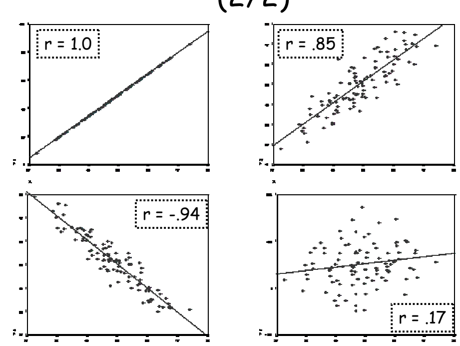
  = 0.9998
- Koefisien determinasi:

  $r^2 = (0.9998)^2 = 0.9996$
- Maka, 99.96% dari variasi dalam waktu untuk membaca file dijelaskan oleh model linier
- Ingat, korelasi bukanlah sebab-akibat!
  - File besar mungkin memang menyebabkan lebih banyak waktu untuk read
  - Tetapi, sebagai contoh, waktu tidak menyebabkan message menjadi lebih lama

Contoh Visual Korelasi (1/2)



Contoh Visual Korelasi (2/2)



Multiple Linear Regression (1/2)

- Termasuk efek dari beberapa variabel input yang secar linier berelasi dengan satu output
- Ekstensi langsung dari regrasi tunggal
- Pertama, pertimbangkan 2 variabel. Perlu:

$$y = b_0 + b_1x_1 + b_2x_2$$

- Buat n pengukuran (x_{1i}, x_{2i}, y_i) dan:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + e_i$$

- Sama seperti sebelumnya, ingin meminimalkan sum square dari error residual (e_i):

$$SSE = \Sigma e_i^2 = \Sigma(y_i-b_0-b_1x_{1i}-b_2x_{2i})^2$$

Multiple Linear Regression (2/2)

- *Sama seperti sebelumnya, minimal jika derivatif parsial 0*

$$nb_0 + b_1\Sigma x_{1i} + b_2\Sigma x_{2i} = \Sigma y_i$$
$$b_0\Sigma x_{1i} + b_1\Sigma x_{1i}^2 + b_2\Sigma x_{1i}x_{2i} = \Sigma x_{1i}y_i$$
$$b_0\Sigma x_{2i} + b_1\Sigma x_{1i}x_{2i} + b_2\Sigma x_{2i}^2 = \Sigma x_{2i}y_i$$

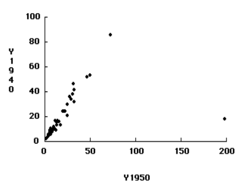
- Tiga persamaan dalam tiga unknowns (b_0, b_1, b_2)
  - Pecahkan menggunakan berbagai varietas software
- Generalisasi:

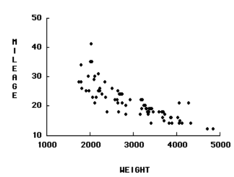
$$y = b_0 + b_1x_1 + \dots + b_kx_k$$

- Dapat merepresentasikan persamaan sbg matriks dan memecahkannya dengan software yang tersedia

Memverivikasi Linieritas (1/2)

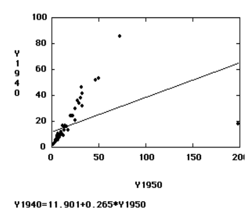
- Dilakukan dengan uji visual sebelum regresi

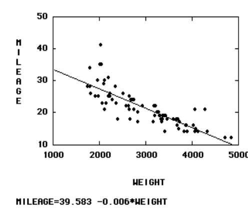




Memverivikasi Linieritas (2/2)

- Regresi Linier mungkin bukan model yang terbaik





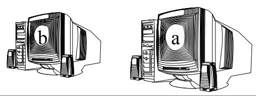
Outline

- Introduksi (sudah)
- Dasar-dasar (sudah)
- Indeks Tendensi Sentral (sudah)
- Indeks Dispersi (sudah)
- Membandingkan Sistem (sudah)
- Misc (sudah)
- Regresi (sudah)
- ANOVA (berikut)

Analysis of Variance (ANOVA)

- Mempartisi variasi kedalam bagian yang dapat dijelaskan dan bagian yang tidak dapat
- Contoh:
  - Mudah melihat regresi yang menjelaskan 70% variasi tidak sebaik satu yang dapat menjelaskan 90% variasi
  - Tetapi berapa banyak dari variasi yang menjelaskan adalah baik?
- Masuk: ANOVA

Perbandingan Sebelum-dan-Sesudah



Pengukuran (i)	Sebelum (b_i)	Sesudah (a_i)	Selisih ($d_i = b_i - a_i$)
1	85	86	-1
2	83	88	-5
3	94	90	4
4	90	95	-5
5	88	91	-3
6	87	83	4

Mean selisih $\underline{d}$ = -1, Deviasi standar s_d = 4.15

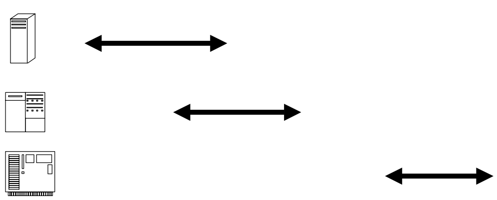
Perbandingan Sebelum-dan-Sesudah

Mean selisih $\underline{d}$ = -1
Deviasi standar s_d = 4.15

- Dari mean selisih, tampak bahwa perubahan sistem mereduksi kinerja
- Akan tetapi, deviasi standar besar
- Apakah variasi antara dua sistem (alternatif) lebih besar dari variasi (error) dalam pengukuran?
- Confidence intervals dapat berlaku, tetapi bagaimana bila lebih dari dua alternatif?

Membandingkan Lebih dari Dua Alternatif

- Pendekatan Naif
  - Membandingkan confidence intervals



- Perlu melakukan untuk tiap pasangan. Tumbuh pesat.
- Mis- 7 alternatif membutuhkan 21 pasang perbandingan
- [(7 pilih 2) = (7)(6) / (2)(1) = 42]
- Ditambah, bukan kejutan untuk menemukan 1 pasang berbeda (pada 95%)

ANOVA – Analysis of Variance (1/2)

- Memisahkan total variasi terobservasi dalam set pengukuran kedalam:
  - (1) Variasi dalam satu sistem
    - Karena error pengukuran tak terkontrol
  - (2) Variasi antar sistem
    - Karena real differences + random error
- Apakah variasi (2) secara statistik lebih besar dari variasi (1)?

ANOVA – Analysis of Variance (2/2)

- Buat n pengukuran k alternatif
- y_{ij} = pengukuran ke i pada alternatif ke j
- Asumsi, error adalah:
  - Independen
  - Terdistribusi Normal

(Contoh panjang, berikut)

Semua Pengukuran untuk Semua Alternatif

	Alternatif					
Pengu-kuran	1	2	…	j	…	k
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Mean Kolom	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Efek	α_1	α_2	…	α_j	…	α_k

Mean Kolom

- Mean kolom adalah nilai rata-rata semua pengukuran dalam satu alternatif
  - Kinerja rata-rata dari stau alternatif

$$\bar{y}_{.j} = \frac{\sum_{i=1}^{n} y_{ij}}{n}$$

	Alternatif					
Pengu-kuran	1	2	...	j	...	k
1	y_{11}	y_{12}	...	y_{1j}	...	y_{k1}
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2k}
...	...	...	...	...	...	...
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ik}
...	...	...	...	...	...	...
n	y_{n1}	y_{n2}	...	y_{nj}	...	y_{nk}
Mean Kolom	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	...	$\bar{y}_{.k}$
Efek	α_1	α_2	...	α_j	...	α_k

Error = Deviasi dari Mean Kolom

- $y_{ij} = y_j + e_{ij}$
- Dimana e_{ij} = error dlm pengukuran

	Alternatif					
Pengu-kuran	1	2	...	j	...	k
1	y_{11}	y_{12}	...	y_{1j}	...	y_{k1}
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2k}
...	...	...	...	...	...	...
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ik}
...	...	...	...	...	...	...
n	y_{n1}	y_{n2}	...	y_{nj}	...	y_{nk}
Mean kolom	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	...	$\bar{y}_{.k}$
Efek	α_1	α_2	...	α_j	...	α_k

Mean Keseluruhan

- Rata-rata semua pengukuran dari seluruh alternatif

$$\bar{y}_{..} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} y_{ij}}{kn}$$

	Alternatif					
Pengu-kuran	1	2	...	j	...	k
1	y_{11}	y_{12}	...	y_{1j}	...	y_{k1}
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2k}
...	...	...	...	...	...	...
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ik}
...	...	...	...	...	...	...
n	y_{n1}	y_{n2}	...	y_{nj}	...	y_{nk}
Mean kolom	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	...	$\bar{y}_{.k}$
Efek	α_1	α_2	...	α_j	...	α_k

Efek = Deviasi dari Mean Keseluruhan

- $y_j = \bar{y} + \alpha_j$
- α_j = deviasi mean kolom dari mean keseluruhan
  = effek dari alternatif j

	Alternatif					
Pengu-kuran	1	2	...	j	...	k
1	y_{11}	y_{12}	...	y_{1j}	...	y_{k1}
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2k}
...	...	...	...	...	...	...
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ik}
...	...	...	...	...	...	...
n	y_{n1}	y_{n2}	...	y_{nj}	...	y_{nk}
Mean kol.	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	...	$\bar{y}_{.k}$
Efek	α_1	α_2	...	α_j	...	α_k

Efek dan Error

- *Efek* adalah jarak dari mean keseluruhan
  - Secara horizon lintas alternatif
- *Error* jarak dari mean kolom
  - Secara vertikal dalam satu alternatif
  - Juga, error lintas alternatif
- Pengukuran individual kemudian adalah:

$$y_{ij} = \bar{y}_{..} + \alpha_j + e_{ij}$$

Sum of Squares of Differences

- ***SST*** = selisih antara tiap pengukuran dan man keseluruhan
- ***SSA*** = variasi karena efek dari alternatif
- ***SSE*** = variasi karena errors dalam pengukuran

$$SSA = n\sum_{j=1}^{k}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{.j}\right)^2$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2$$

$$\boxed{SST = SSA + SSE}$$

ANOVA

- Memisahkan variasi dalam nilai terukur kedalam:
  1. Variasi karena efek dari alternatif
     - SSA – variasi lintas kolom
  2. Variasi karena errors
     - SSE – variasi dalam satu kolom
- Jika selisih antar alternatif karena selisih real:

  → SSA secara secara statistik lebih besar dari SSE

Membandingkan SSE dan SSA

- Pendekatan sederhana
  - *SSA / SST* = fraksi total variasi terjelaskan oleh selisih antar alternatif
  - *SSE / SST* = fraksi total variasi karena error eksperimental
- Apakah ini secara statistik signifikan?
- Variance = mean square values

  = total variasi / derajat kebebasan

  s_x^2 = SSx / df(SSx)
- (Derajat kebebasan adalah jumlah term independen dalam sum)

Derajat Kebebasan untuk Efek

- *df*(*SSA*) = *k* – 1, karena *k* alternatif

	Alternatif					
Pengu-kuran	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
I	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
N	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Mean kolom	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Efek	α_1	α_2		α_j		α_k

Derajat Kebebasan untuk Errors

- *df*(*SSE*) = *k*(*n* – 1), karena *k* alternatif, masing-masing dgn (*n* – 1) *df*

	Alternatif					
Pengu-kuran	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Mean kolom	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Efek	α_1	α_2	…	α_j	…	α_k

Derajat Kebebasan untuk Total

- *df*(*SST*) = *df*(*SSA*) + *df*(*SSE*) = *kn* - 1

	Alternatif					
Pengu-kuran	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Mean kolom	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Efek	α_1	α_2	…	α_j	…	α_k

Variances dari Sum of Squares (Mean Square Value)

$$s_a^2 = \frac{SSA}{k-1}$$

$$s_e^2 = \frac{SSE}{k(n-1)}$$

Membandingkan Variances

- Menggunakan F-test untuk membandingkan rasio variance
  - F-test digunakan untuk menguji jika deviasi standar dari populasi adalah sama.

$$F = \frac{s_a^2}{s_e^2}$$

$$F_{[1-\alpha;df(num),df(denom)]} = \text{tabulated critical values}$$

- Jika $F_{computed} > F_{table}$ *untuk* α
  → Kita memiliki $(1 - \alpha) * 100\%$ confidence bahwa variasi karena selisih aktual dalam alternatif, SSA, secara statistik lebih besar dari variasi karena errors, SSE.

Ringkasan ANOVA

Variation	Alternatives	Error	Total
Sum of squares	SSA	SSE	SST
Deg freedom	$k-1$	$k(n-1)$	$kn-1$
Mean square	$s_a^2 = SSA/(k-1)$	$s_e^2 = SSE/[k(n-1)]$	
Computed F	s_a^2/s_e^2		
Tabulated F	$F_{[1-\alpha;(k-1),k(n-1)]}$		

(Contoh, berikut)

Contoh ANOVA (1/2)

	Alternatif			
Pengukuran	**1**	**2**	**3**	**Mean keseluruhan**
1	0.0972	0.1382	0.7966	
2	0.0971	0.1432	0.5300	
3	0.0969	0.1382	0.5152	
4	0.1954	0.1730	0.6675	
5	0.0974	0.1383	0.5298	
Mean kolom	0.1168	0.1462	0.6078	0.2903
Efek	-0.1735	-0.1441	0.3175	

Contoh ANOVA (2/2)

Variation	Alternatives	Error	Total
Sum of squares	$SSA = 0.7585$	$SSE = 0.0685$	$SST = 0.8270$
Deg freedom	$k-1=2$	$k(n-1)=12$	$kn-1=14$
Mean square	$s_a^2 = 0.3793$	$s_e^2 = 0.0057$	
Computed F	$0.3793/0.0057 = 66.4$		
Tabulated F	$F_{[0.95;2,12]} = 3.89$		

- SSA/SST = 0.7585/0.8270 = 0.917
  → 91.7% dari total variasi dalam pengukuran karena selisih antar alternatif
- SSE/SST = 0.0685/0.8270 = 0.083
  → 8.3% dari total variasi dalam pengukuran karena noise dalam pengukuran
- Computed *F* statistic > tabulated *F* statistic
  → 95% confidence bahwa perbedaan antar alternatif adalah secara statistik signifikan.

Ringkasan ANOVA

- Berguna untuk mempartisi total variasi kedalam komponen
  - Error eksperimental
  - Variasi antar alternatif
- Membandingkan lebih dari satu alternatif
- Ingat, tidak mengatakan *dimana* perbedaan itu berada
  - Gunakan confidence intervals untuk pasangan
  - Atau gunakan contrast