Statistika Non Parametrik dan Penerapannya dalam Penelitian Manajemen

Bambang Suryoatmono

Bagian 1

Analisis Regresi Sederhana
(*Simple Regression Analysis*)

Pendahuluan

- Analisis Regresi: proses membuat fungsi atau model matematis yang dapat digunakan untuk memprediksi atau menentukan satu variabel dari variabel lainnya.
- Regresi Sederhana (*bivariate linear regression*): regresi yang hanya melibatkan dua variabel.
  - Variabel bergantung (*dependent variable*): variabel yang akan diprediksi (y)
  - Variabel bebas (*explanatory variable = independent variable):* prediktor
  - Hanya hubungan linear antara kedua variabel
- Hubungan non linear dan model regresi dengan lebih dari satu variabel bebas: model regresi berganda (*multiple regression model)*

Model-model Regresi

- Model Deterministik

$$y = \beta_0 + \beta_1 x$$

- Model Probabilistik

$$y = \beta_0 + \beta_1 x + \varepsilon$$

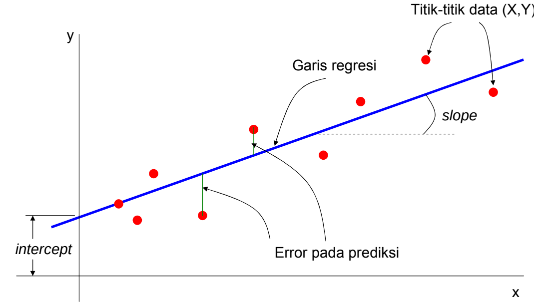
- β_0 = *intercept* populasi
- β_1 = kemiringan (*slope*) populasi

Pers. Garis Regresi Sederhana

$$\hat{y} = b_0 + b_1 x$$

- b_0 = *intercept* sampel
- b_1 = *slope* sampel
- Keduanya dicari dengan analisis kuadrat terkecil (*least square analysis*): proses di mana model regresi dicari yang menghasilkan jumlah error kuadrat terkecil

Error pada prediksi



Slope dan *Intercept* Sampel

$$SS_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$SS_{xx} = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

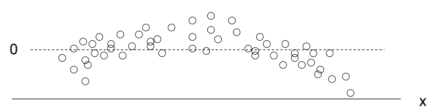
$$b_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$b_0 = \bar{y} - b_1\bar{x} = \frac{\Sigma y}{n} - b_1\frac{\Sigma x}{n}$$

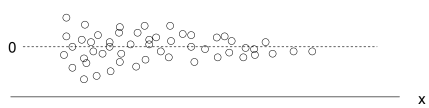
Analisis Residual

- Residual = error garis regresi = perbedaan antara y prediksi (dari persamaan regresi) dan y aktual = $y - \hat{y}$
- Tujuan analisis Residual: menguji sebagian atau seluruh asumsi yang mendasari regresi sederhana, yaitu:
  - Model adalah linear
  - Suku error mempunyai varians yang konstan
  - Semua suku error: independen
  - Suku error terdistribusi normal

Residual Plot

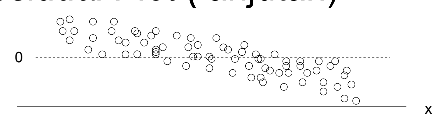


Nonlinear Residual Plot

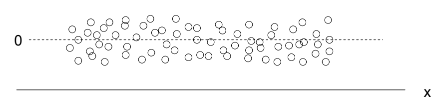


Nonconstant Error Variance

Residual Plot (lanjutan)



Nonindependent Error Terms



Healthy Residual Graph

Sum of Squares of Error (SSE)

- Cara alternatif untuk mempelajari error pada regresi
- Merupakan satu ukuran error pada regresi

$$SSE = \Sigma(y-\hat{y})^2 = \Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$$

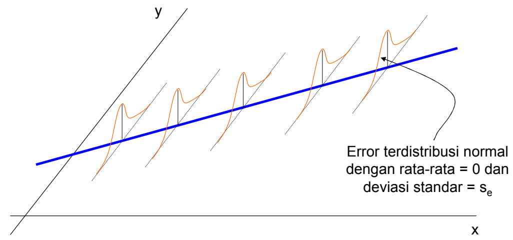
Standard Error of The Estimate s_e

- s_e adalah deviasi standar error pada model regresi

$$s_e = \sqrt{\frac{SSE}{n-2}}$$

- Dapat digunakan untuk
  - mempelajari error pada model
  - mengestimasi *outliers*

Standard Error of The Estimate s_e (lanjutan)



Koefisien Determinasi r^2

- r^2 = variabilitas variabel bergantung yang diakibatkan oleh variabel bebas x
- Bernilai antara 0 sampai dengan 1
- r^2 = 0 artinya: prediktor (x) tidak mempengaruhi variabilitas y;
- r^2 = 1 artinya: variabilitas y seluruhnya diakibatkan oleh prediktor x

Koefisien Determinasi r^2 (lanjutan)

$$SS_{yy} = \Sigma(y-\bar{y})^2 = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$SS_{yy} = \underbrace{SSR}_{\text{regresi}} + \underbrace{SSE}_{\text{error}}$$

$$r^2 = \frac{SSR}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}} \text{ atau lebih mudah dihitung dengan}$$

$$r^2 = \frac{b_1^2 SS_{xx}}{SS_{yy}}$$

$$0 \le r^2 \le 1$$

Koefisien Korelasi Pearson

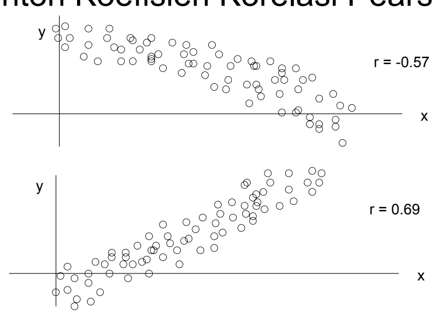
- Korelasi = derajat keterkaitan antara dua variabel

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2(y-\bar{y})^2}}$$

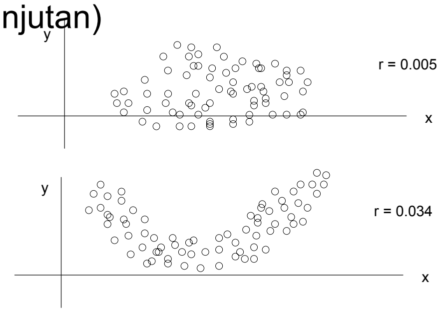
$$-1 \le r \le 1$$

- r = 0 → tidak ada hubungan linear antara kedua variabel
- r = 1 → ada korelasi positif sempurna antara kedua variabel
- r = -1 → ada korelasi negatif sempurna antara kedua variabel

Contoh Koefisien Korelasi Pearson

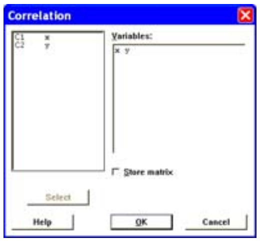


Contoh Koefisien Korelasi Pearson (lanjutan)



Koefisien Korelasi Pearson r dengan MINITAB

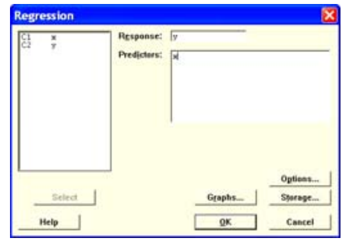
- Stat → Basic Statistics → Correlation

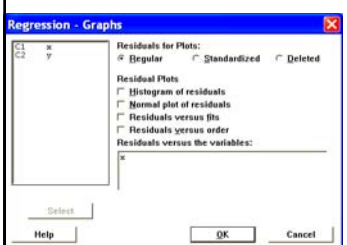


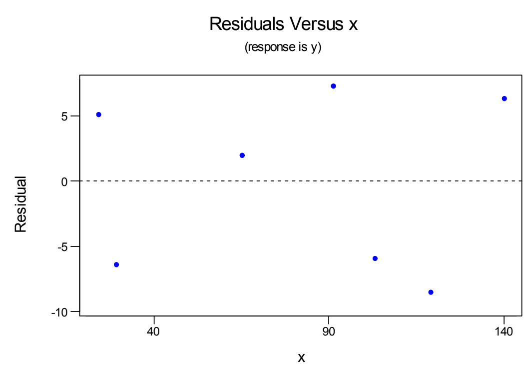
Analisis Regresi dengan MINITAB

Row	x	y
1	140	25
2	119	29
3	103	46
4	91	70
5	65	88
6	29	112
7	24	128

- Stat → Regression→ Regression







Regression Analysis

p-value untuk menguji *slope*

```
The regression equation is
y = 144 - 0.898 x

Predictor      Coef      StDev          T       P
Constant    144.414      6.220      23.22   0.000
x          -0.89824    0.06816     -13.18   0.000

S = 7.377       R-Sq = 97.2%     R-Sq(adj) = 96.6%

Analysis of Variance

Source      DF        SS        MS        F        P
Regression   1    9452.7    9452.7   173.69    0.000
Error        5     272.1      54.4
Total        6    9724.9
```

$s_e = \sqrt{MSE}$

p-value untuk menguji *overall model*

Testing the *Slope*

- Statistik uji: $t = \dfrac{b_1 - \beta_{1,0}}{s_b}$ dengan

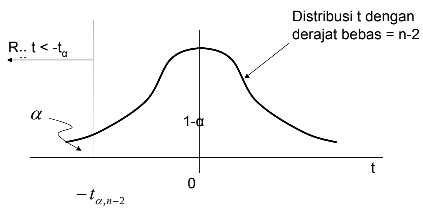
$$s_b = \frac{s_e}{\sqrt{SS_{xx}}}$$

$$s_e = \sqrt{\frac{SSE}{n-2}}$$

$$SS_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

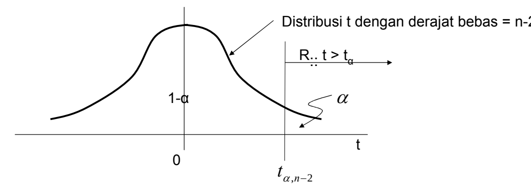
Testing the *Slope* (lanjutan)

- H_0: $\beta_1 = \beta_{1,0}$ vs H_a: $\beta_1 < \beta_{1,0}$



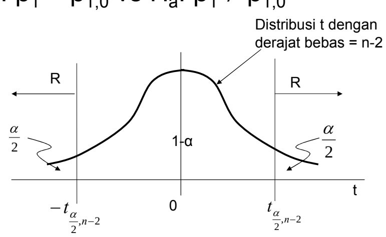
Testing the *Slope* (lanjutan)

- H_0: $\beta_1 = \beta_{1,0}$ vs H_a: $\beta_1 > \beta_{1,0}$



Testing the *Slope* (lanjutan)

- H_0: $\beta_1 = \beta_{1,0}$ vs H_a: $\beta_1 \neq \beta_{1,0}$



Catatan: cara p-value juga dapat digunakan. Tolak H_0 jika p-value < α

Testing the Overall Model (Uji F) Tabel ANOVA

Source	DF	SS	MS	F
Regresi	k	SSR	$\mathrm{MSR} = \frac{\mathrm{SSR}}{k}$	$\mathrm{F} = \frac{\mathrm{MSR}}{\mathrm{MSE}}$
Residual Error	n – k – 1	SSE	$\mathrm{MSE} = \frac{\mathrm{SSE}}{\mathrm{n-k-1}}$	
Jumlah	n – 1	SS_{yy}		

Catatan:
- k = banyak variabel bebas (untuk regresi sederhana, k = 1)
- Derajat bebas F adalah k (pembilang) dan N-k-1 (penyebut)

Estimasi

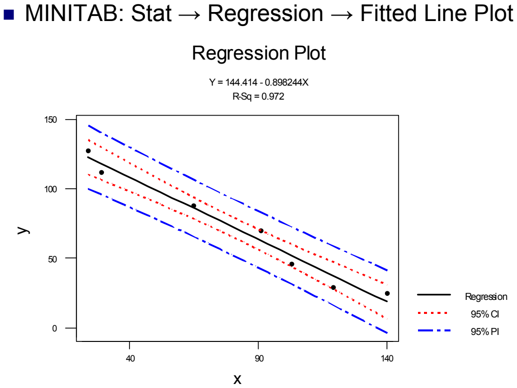
- CI untuk mengestimasi Rata-rata Bersyarat untuk y: $\mu_{y|x}$ untuk harga x yang ditetapkan

$$\hat{y} \pm t_{\frac{\alpha}{2},n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}$$

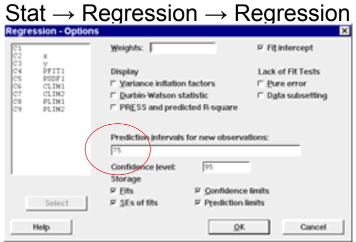
- Interval Prediksi (PI) untuk Mengestimasi Harga Tunggal y untuk harga x yang ditetapkan

$$\hat{y} \pm t_{\frac{\alpha}{2},n-2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}$$

- MINITAB: Stat → Regression → Fitted Line Plot



- MINITAB: Stat → Regression → Regression → Option




```
Predicted Values for New Observations

New Obs     Fit     SE Fit         95.0% CI             95.0% PI
1         77.05       2.82   (   69.79,   84.31)  (   56.74,   97.35)

Values of Predictors for New Observations

New Obs           x
1              75.0
```

Bagian 2

Analisis Regresi Berganda

Analisis Regresi Berganda

- adalah analisis regresi dengan dua atau lebih variabel bebas atau dengan sedikitnya satu prediktor non linear
- Model regresi berganda probabilistik:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots \beta_k x_k + \varepsilon$$

  - k = banyaknya variabel bebas
  - β_0 = konstanta regresi
  - β_i = koefieisn regresi parsial untuk variabel independen I; menunjukkan bertambahnya y apabila variabel independen I meningkat 1 unit dan variabel independen lainnya tidak berubah
  - x_2 dapat berupa x_1^2 (suku non linear dari x_1)

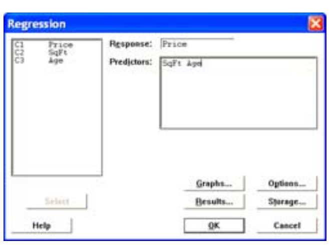
Estimasi y

- Estimasi y dengan menggunakan informasi dari sampel

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots\ldots + b_k x_k$$

  - $\hat{y}$ = nilai y prediksi
  - b_0 = estimasi konstanta regresi
  - b_i = estimasi koefisien regresi 1
- MINITAB: Stat → Regression→ Regression

Row	Price	SqFt	Age
1	63.0	1605	35
2	65.1	2489	45
3	69.9	1553	20
4	76.8	2404	32
5	73.9	1884	25
6	77.9	1558	14
7	74.9	1748	8
8	78.0	3105	10
9	79.0	1682	28
10	83.4	2470	30
11	79.5	1820	2
12	83.9	2143	6
13	79.7	2121	14
14	84.5	2485	9
15	96.0	2300	19
16	109.5	2714	4
17	102.5	2463	5
18	121.0	3076	7
19	104.9	3048	3
20	128.0	3267	6
21	129.0	3069	10
22	117.9	4765	11
23	140.0	4540	8




```
Regression Analysis: Price versus SqFt, Age

The regression equation is
Price = 57.4 + 0.0177 SqFt - 0.666 Age

Predictor        Coef     SE Coef          T        P
Constant        57.35       10.01       5.73    0.000
SqFt         0.017718    0.003146       5.63    0.000
Age           -0.6663      0.2280      -2.92    0.008

S = 11.96       R-Sq = 74.1%     R-Sq(adj) = 71.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      8189.7      4094.9     28.63    0.000
Residual Error    20      2861.0       143.1
Total             22     11050.7

Source       DF      Seq SS
SqFt          1      6967.8
Age           1      1221.9

Unusual Observations
Obs       SqFt      Price        Fit      SE Fit    Residual    St Resid
  8       3105      78.00     105.70        3.08      -27.70      -2.40R
 21       3069     129.00     105.06        3.03       23.94       2.07R

R denotes an observation with a large standardized residual
```

Menguji *Overall Model*

- H_0: $\beta_1 = \beta_2 = ... = \beta_k = 0$
- H_a: sedikitnya satu koefisien regresi ≠ 0
- Statistik uji: F (lihat tabel ANOVA)

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

- Pada contoh di atas: nilai p (=0.000) < α (= 5%) → tolak H_0. Jadi, sedikitnya satu koefisien regresi ≠ 0

Menguji Signifikansi Koefisien Regresi

- H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$
  Pada contoh di atas, nilai p untuk β_1 adalah 0.000 < α (= 5%) → tolak H_0. Artinya, variabel SqFt berpengaruh secara signifikan terhadap variabel Price.
- H_0: $\beta_2 = 0$ versus H_a: $\beta_2 \neq 0$
  Pada contoh di atas, nilai p untuk β_2 adalah 0.008 < α (= 5%) → tolak H_0. Artinya, variabel Age berpengaruh secara signifikan terhadap variabel Price.

Residual, SSE, *Standard Error of the Estimate,* dan R^2

- Residual = $y - \hat{y}$
- $SSE = \Sigma(y - \hat{y})^2$
- *Standard Error of the Estimate* $s_e = \sqrt{\frac{SSE}{n-k-1}}$
- Koefisien Determinasi Berganda $R^2 = 1 - \frac{SSE}{SS_{yy}}$

R^2 *adjusted*

- R^2 selalu membesar (atau setidaknya tetap) apabila variabel bebas ditambahkan
- Untuk memperhitungkan
  - informasi tambahan pada regresi setiap kali variabel independen ditambahkan, dan
  - Perubahan derajat bebas pada regresi,

  dibuatlah R^2 yang disesuaikan:

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SS_{yy}/(n-1)}$$

Bagian 3

Membangun Model Regresi Berganda

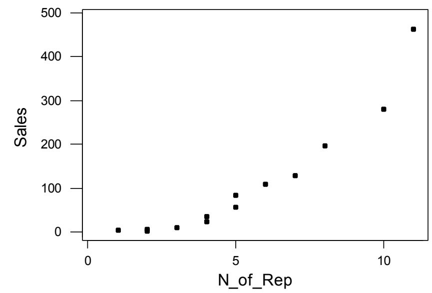
Model Regresi Polinomial

- adalah model regresi yang merupakan model orde dua atau lebih.
- Model kuadratik adalah model regresi berganda di mana prediktornya adalah satu variabel dan kuadrat dari variabel tersebut.

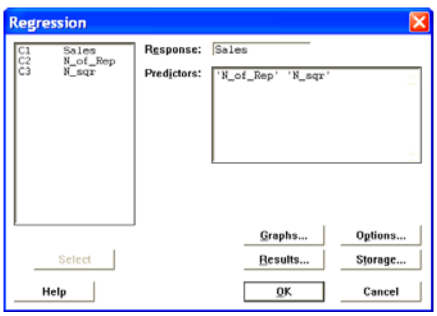
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon$$

Row	Sales	N_of_Rep	N_sqr
1	2.1	2	4
2	3.6	1	1
3	6.2	2	4
4	10.4	3	9
5	22.8	4	16
6	35.6	4	16
7	57.1	5	25
8	83.5	5	25
9	109.4	6	36
10	128.6	7	49
11	196.8	8	64
12	280.0	10	100
13	462.3	11	121

dikuadratkan



- MINITAB: Stat → Regression → Regression



Model Kuadratik

Regression Analysis: Sales versus N_of_Rep, N_sqr

The regression equation is
Sales = 18.1 - 15.7 N_of_Rep + 4.75 N_sqr

Predictor	Coef	SE Coef	T	P
Constant	18.07	24.67	0.73	0.481
N_of_Rep	-15.723	9.550	-1.65	0.131
N_sqr	4.7504	0.7759	6.12	0.000

S = 24.59   R-Sq = 97.3%   R-Sq(adj) = 96.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	215069	107534	177.79	0.000
Residual Error	10	6048	605		
Total	12	221117			

Model Linear

Regression Analysis: Sales versus N_of_Rep

The regression equation is
Sales = - 107 + 41.0 N_of_Rep

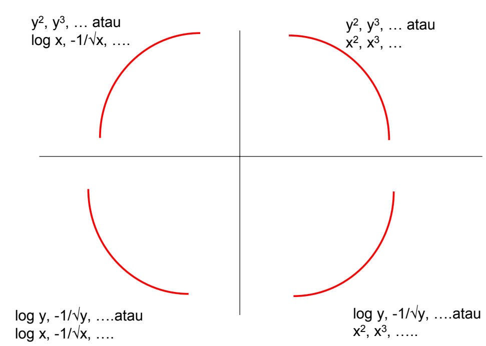
Predictor	Coef	SE Coef	T	P
Constant	-107.03	28.74	-3.72	0.003
N_of_Rep	41.026	4.779	8.58	0.000

S = 51.10   R-Sq = 87.0%   R-Sq(adj) = 85.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	192395	192395	73.69	0.000
Residual Error	11	28721	2611		
Total	12	221117			

Transformasi Tukey



Model Regresi dengan Interaksi

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

suku interaksi

- x_1x_2 adalah suku interaksi
- Di dalam proses regresi, x_1x_2 disubstitusi dengan variabel x_3 sehingga model regresinya menjadi

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Transformasi Model

- Contoh: $y = \beta_0 x^{\beta_1} \varepsilon$

jelas bukan merupakan model linear. Namun jika ditransformasi menjadi

$$\log y = \log \beta_0 + \beta_1 \log x + \varepsilon$$

$$y' = \beta_0' + \beta_1' x' \text{ dengan}$$

$$y' = \log\ y$$

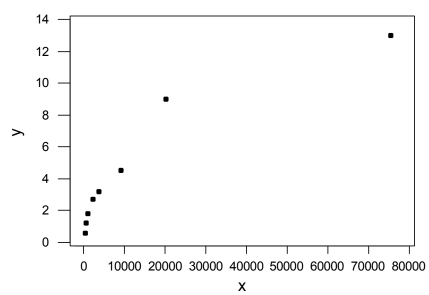
$$\beta_0' = \log \beta_0 \text{ dan}$$

$$x' = \log\ x$$

Contoh Data

Row	y	x	log_y	log_x
1	1.2	450	0.07918	2.65321
2	9.0	20200	0.95424	4.30535
3	4.5	9060	0.65321	3.95713
4	3.2	3500	0.50515	3.54407
5	13.0	75600	1.11394	4.87852
6	0.6	175	-0.22185	2.24304
7	1.8	800	0.25527	2.90309
8	2.7	2100	0.43136	3.32222

Plot x versus y



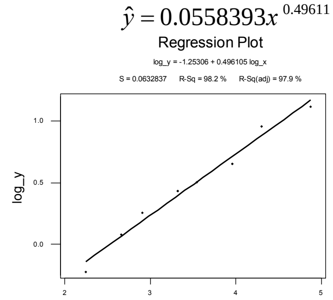
Output MINITAB

```
Regression Analysis: log_y versus log_x

The regression equation is
log_y = - 1.25 + 0.496 log_x

Predictor        Coef     SE Coef          T        P
Constant     -1.25306     0.09693     -12.93    0.000
log_x         0.49611     0.02713      18.28    0.000

S = 0.06328     R-Sq = 98.2%     R-Sq(adj) = 97.9%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      1.3389      1.3389    334.32    0.000
Residual Error     6      0.0240      0.0040
Total              7      1.3629
```

$$b_o = 10^{-1.25306} = 0.0558393$$

Jadi, model regresi dinyatakan dalam variabel asal adalah

$$\hat{y} = 0.0558393 x^{0.49611}$$



Variabel Indikator (*dummy*)

- Variabel kualitatif hanya memberikan informasi data pada level nominal atau ordinal
- Variabel ini disebut juga dengan variabel *dummy* atau variabel indikator
- Jika variabel indikator mempunyai c kategori, maka dibutuhkan c-1 variabel *dummy*

Contoh Variabel Indikator

- Variabel Kualitatif: Lokasi tempat tinggal. Ada 4 pilihan: Jakarta, Bandung, Surabaya, Medan (4 kategori).
- Jadi butuh 3 variabel *dummy*. Sebut saja: Jakarta, Bandung, Surabaya.

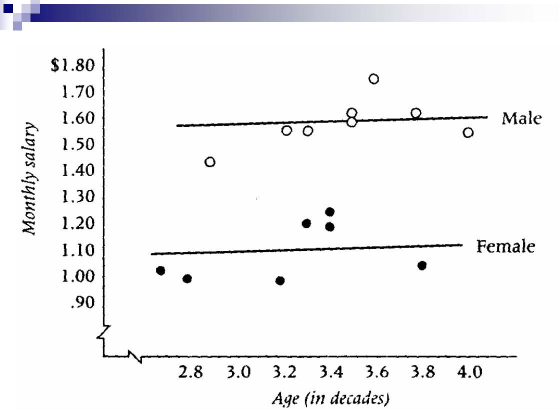
Tempat tinggal di	Variabel Dummy		
	Jkt	Bdg	Sby
Jkt	1	0	0
Bdg	0	1	0
Sby	0	0	1
Mdn	0	0	0

Contoh

Row	Salary	Age	Gender
1	1.548	3.2	1
2	1.629	3.8	1
3	1.011	2.7	0
4	1.229	3.4	0
5	1.746	3.6	1
6	1.528	4.1	1
7	1.018	3.8	0
8	1.190	3.4	0
9	1.551	3.3	1
10	0.985	3.2	0
11	1.610	3.5	1
12	1.432	2.9	1
13	1.215	3.3	0
14	0.990	2.8	0
15	1.585	3.5	1

Gender: 1 = male, 0 = female

```
The regression equation is
Salary = 0.732 + 0.111 Age + 0.459 Gender

Predictor        Coef     SE Coef         T        P
Constant       0.7321      0.2356      3.11    0.009
Age           0.11122     0.07208      1.54    0.149
Gender        0.45868     0.05346      8.58    0.000

S = 0.09679     R-Sq = 89.0%     R-Sq(adj) = 87.2%

Analysis of Variance

Source           DF          SS          MS         F        P
Regression        2     0.90949     0.45474     48.54    0.000
Residual Error   12     0.11242     0.00937
Total            14     1.02191
```



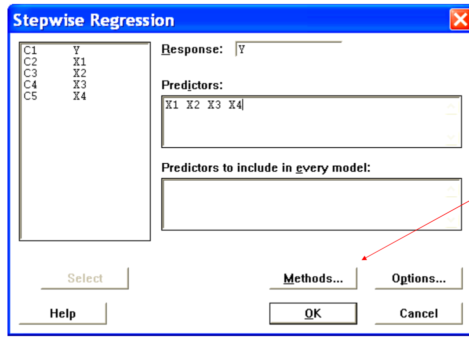
Pembentukan model: Prosedur Pencarian

- Problem: Misalkan ada 3 variabel bebas yang berpotensi mempengaruhi 1 variabel bergantung.
- Prosedur Pencarian adalah proses di mana lebih dari satu model regresi berganda dikembangkan untuk satu basis data, dan model-model tersebut dibandingkan dan disortir berdasarkan kriteria yang bergantung pada prosedur yang digunakan:
  - *All Possible Regression*
  - *Stepwise Regression*
  - *Forward Selection*
  - *Backward Selection*

Contoh Data

Row	Y	X1	X2	X3	X4
1	101	2	77	1.2	42
2	127	4	72	1.7	26
3	98	9	69	2.4	47
4	79	5	53	2.6	65
5	118	3	88	2.9	37
6	114	1	53	2.7	28
7	110	3	82	2.8	29
8	94	2	61	2.6	22
9	96	8	60	2.4	48
10	73	6	64	2.1	42
11	108	2	76	1.8	34
12	124	5	74	2.2	11
13	82	6	50	1.5	61
14	89	9	57	1.6	53
15	76	1	72	2.0	72
16	109	3	74	2.8	36
17	123	2	99	2.6	17
18	125	6	81	2.5	48

- MINITAB: Stat → Regression→ Stepwise



Untuk memilih Stepwise, Forward, atau Backward

```
Stepwise Regression: Y versus X1, X2, X3, X4

  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

 Response is     Y     on  4 predictors, with N =   18

    Step             1        2
Constant        133.53    91.01

X4               -0.78    -0.60
T-Value          -4.20    -3.22
P-Value          0.001    0.006

X2                         0.51
T-Value                    2.15
P-Value                   0.048

S                 12.6     11.4
R-Sq             52.46    63.69
R-Sq(adj)        49.49    58.85
C-p                3.4      1.3
```

Kesimpulan: hanya x_2 dan x_4 yang sebaiknya digunakan dalam model. Variabel x_1 dan x_3 tidak signifikan terhadap perubahan y.

```
Stepwise Regression: y versus x1, x2, x3, x4

Forward selection.  Alpha-to-Enter: 0.1

 Response is     y     on  4 predictors, with N =   18


    Step             1        2
Constant        133.53    91.01

x4               -0.78    -0.60
T-Value          -4.20    -3.22
P-Value          0.001    0.006

x2                         0.51
T-Value                    2.15
P-Value                   0.048

S                 12.6     11.4
R-Sq             52.46    63.69
R-Sq(adj)        49.49    58.85
C-p                3.4      1.3
```

Kesimpulan: hanya x_2 dan x_4 yang sebaiknya digunakan dalam model. Variabel x_1 dan x_3 tidak signifikan terhadap perubahan y.

```
Stepwise Regression: y versus x1, x2, x3, x4

Backward elimination.  Alpha-to-Remove: 0.1

 Response is     y     on  4 predictors, with N =   18


    Step             1        2        3
Constant         83.96    86.93    91.01

x1                 0.6      0.6
T-Value           0.50     0.51
P-Value          0.623    0.617

x2                0.53     0.54     0.51
T-Value           2.04     2.16     2.15
P-Value          0.062    0.049    0.048

x3                 1.4
T-Value           0.23
P-Value          0.824

x4               -0.61    -0.62    -0.60
T-Value          -2.98    -3.18    -3.22
P-Value          0.011    0.007    0.006

S                 12.1     11.7     11.4
R-Sq             64.49    64.35    63.69
R-Sq(adj)        53.57    56.71    58.85
C-p                5.0      3.1      1.3
```

Kesimpulan: hanya x_2 dan x_4 yang sebaiknya digunakan dalam model. Variabel x_1 dan x_3 tidak signifikan terhadap perubahan y.

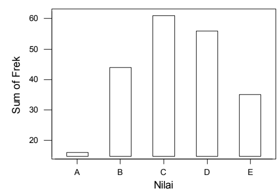
Bagian 4

Analisis Data Kategori:

- Chi-Square Goodness of Fit Test
- Chi-Square Test of Independence

Data Kategori

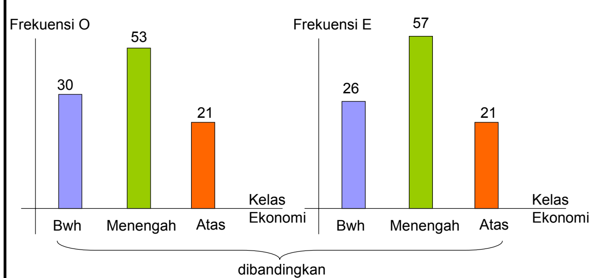
- adalah data non numerik yang merupakan hitungan frekuensi dua atau lebih kategori dari satu atau lebih variabel
- Contoh:



Chi-Square Goodness of Fit Test

- digunakan untuk menganalisis probabilitas *trial* distribusi multinomial pada satu dimensi.
- Contoh: Kelas ekonomi (satu dimensi) dengan kemungkinan *outcome*:
  - Kelas bawah
  - Kelas menengah
  - Kelas atas
- Membandingkan frekuensi kategori teoritis *(expected)* dari populasi, dengan frekuensi kategori aktual *(observed)*, apakah sama atau tidak sama.

Contoh

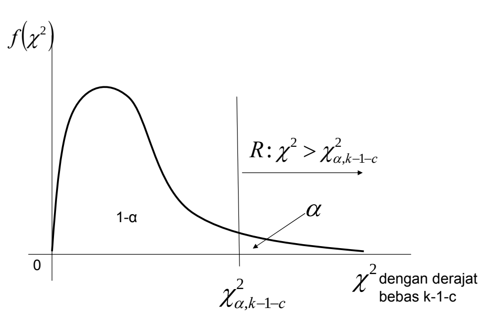


O = *Observed* (yang diamati, aktual)
E = Expected (yang diduga, teoritis)

Uji Hipotesa

- H_0: distribusi yang diamati sama dengan distribusi yang diduga
- H_a: distribusi yang diamati tidak sama dengan distribusi yang diduga
- Statistik uji: $$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$
- df = k – 1 – c
- f_0 = frekuensi hasil pengamatan
- f_e = frekuensi yang diduga
- k= banyaknya kategori
- c = banyaknya parameter yang diestimasi dari data sampel, miaslnya 0 (uniform), 1 (Poisson), 2 (Normal)

Rejection Region R



Contoh Soal

- Di dalam bisnis, kedatangan acak seringkali diasumsikan terdistribusi Poisson. Distribusi ini dicirikan dengan rata-rata kedatangan λ per suatu interval. Misalkan seorang supervisi meyakini bahwa kedatangan acak di suatu bank terdistribusi Poisson dan akan menguji hipotesa ini dengan mengumpulkan informasi. Data berikut ini menunjukkan distribusi frekuensi kedatangan pada interval satu menit di bank tersebut, Gunakan α = 0.05 untuk menentukan apakah kedatangan acak memang terdistribusi Poisson

Data

Banyaknya kedatangan	Frekuensi yang diamati f_o
0	7
1	18
2	25
3	17
4	12
$\geq$5	5

Jawab

- H_0: distribusi yang diamati sama dengan distribusi yang diduga (Poisson)
- H_a: distribusi yang diamati tidak sama dengan distribusi yang diduga (Poisson)
- c = 1 (hanya 1 parameter yang diestimasi, yaitu λ)
- k = 6
- df = k – 1 – c = 6 – 1 – 1 = 4
- α = 5%
- R: $\chi^2 > \chi^2_{0.05,4} = 9.488$

- Estimasi parameter λ

Banyaknya kedatangan	Frekuensi yang diamati f_o	Kedatangan * Frekuensi yang diamati
0	7	0
1	18	18
2	25	50
3	17	51
4	12	48
$\geq$5	5	25
Jumlah	84	192

$$\lambda = \frac{192}{84} = 2.3$$

(rata-rata kedatangan per menit)

- Frekuensi yang diduga

Banyaknya kedatangan	Probabilitas yang diduga (Poisson dengan λ = 2.3)	Frekuensi yang diduga f_e
0	0.1003	8.42
1	0.2306	19.37
2	0.2652	22.28
3	0.2033	17.08
4	0.1169	9.82
$\geq$5	0.0837	7.03
Jumlah		84

- Statistik uji χ^2

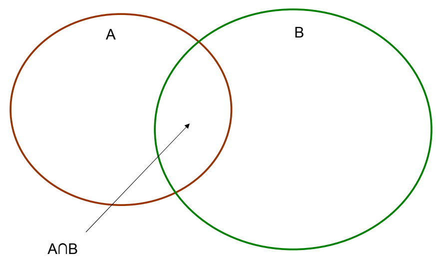
Banyaknya kedatangan	Frekuensi yang diamati f_o	Frekuensi yang diduga f_e	$\frac{(f_o - f_e)^2}{f_e}$
0	7	8.42	0.24
1	118	19.37	0.10
2	25	22.28	0.33
3	17	17.08	0.00
4	12	9.82	0.48
$\geq$5	5	7.03	0.59
Jumlah			$\chi^2 = 1.74$

Karena χ^2 ada di luar R, maka pertahankan H_0. Artinya, memang waktu kedatangan terdistribusi Poisson.

Contingency Analysis: Chi-Square Test of Independence

- digunakan untuk menganalisis frekuensi dua variabel dengan kategori berganda untuk menentukan apakah kedua variabel independen
- Contoh:
- Penghasilan setahun (dalam juta rupiah):
  - a. < 20 juta
  - b. 20 juta sampai dengan 30 juta
  - c. > 30 juta
- Jenis BBM yang biasa digunakan:
  - a. solar
  - b. premium
  - c. premix

Review tentang Probabilitas

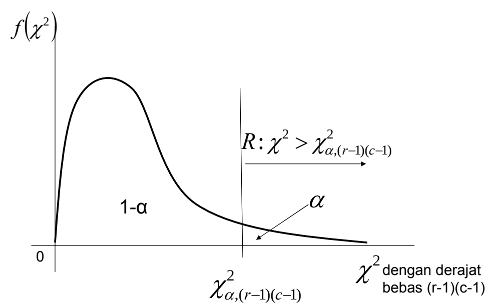


Jika A dan B independen, maka P(A∩B) = P(A) * P(B)
Note: P(A∩B) dapat ditulis P(AB), dibaca Probabilitas (A dan B terjadi)

Uji Hipotesa

- H_0: kedua variabel kategori independen (tidak saling bergantung)
- H_a: kedua variabel kategori saling bergantung
- Statistik uji: $\chi^2 = \sum\sum \frac{(f_o - f_e)^2}{f_e}$
- df = (r – 1)(c – 1)
- r= banyaknya baris
- c = banyaknya kolom
- f_0 = frekuensi hasil pengamatan
- f_e = frekuensi yang diduga $= e_{ij} = \frac{n_i n_j}{N}$
- ni = total baris i
- nj = total kolom j
- N = total semua frekuensi

Rejection Region R



Contoh Soal

- Apakah jenis minuman yang dipesan di sebuah restoran pada saat makan siang tidak bergantung pada usia pemesannya? *Polling* acak pada 309 pemesan minuman pada saat makan siang di restoran ditunjukkan pada tabel berikut. Gunakan α = 0.05 untuk menentukan apakah kedua variabel tidak saling bergantung.

Data

		Minuman yang dipesan		
		Teh/Kopi	Minuman ringan	Lain-lain (susu dll)
Usia	21-34	26	95	18
	35-55	41	40	20
	>55	24	13	32

Jawab

- H_0: jenis minuman yang dipesan tidak bergantung pada usia pemesan
- H_a: jenis minuman yang dipesan bergantung pada usia pemesan
- Statistik uji

$$\chi^2 = \sum\sum \frac{(f_o - f_e)^2}{f_e}$$

- r = 3
- c = 3
- df = (3-1)(3-1) = 4
- α = 5%
- R: $\chi^2 > \chi^2_{0.05,4} = 9.4877$

- Menghitung frekuensi yang diduga f_e

		Minuman yang dipesan			
		Teh/Kopi	Minuman ringan	Lain-lain (susu dll)	
Usia	21-34	(40.94) 26	(66.58) 95	(31.49) 18	139
	35-55	(29.74) 41	(48.38) 40	(22.88) 20	101
	>55	(20.32) 24	(33.05) 13	(15.63) 32	69
		91	148	70	309

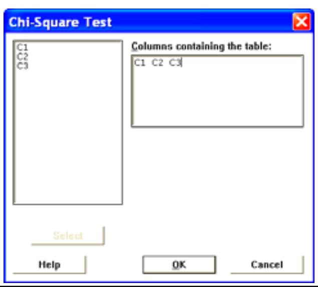
$$e_{12} = \frac{139*148}{309} = 66.58 \qquad e_{31} = \frac{69*91}{309} = 20.32$$

- Statistik uji

$$\chi^2 = \sum\sum \frac{(f_o - f_e)^2}{f_e} = \frac{(26-40.94)^2}{40.94} + \frac{(95-66.58)^2}{66.58} + ... + \frac{(32-15.63)^2}{15.63}$$
$$= 59.41$$

- Karena $\chi^2 > 9.4877$ maka H_0 ditolak. Artinya, jenis minuman yang dipesan pada saat makan siang di suatu restoran bergantung pada usia pemesannya.
- Dengan MINITAB: Stat → Table → Chi-Square Test

Row	C1	C2	C3
1	26	95	18
2	41	40	20
3	24	13	32




```
Chi-Square Test: C1, C2, C3

Expected counts are printed below observed counts

            C1       C2       C3    Total
    1       26       95       18      139
         40.94    66.58    31.49

    2       41       40       20      101
         29.74    48.38    22.88

    3       24       13       32       69
         20.32    33.05    15.63

Total       91      148       70      309

Chi-Sq =  5.449 + 12.135 +  5.778 +
          4.259 +  1.450 +  0.363 +
          0.666 + 12.162 + 17.142 = 59.405
DF = 4, P-Value = 0.000
```

sama dengan yang telah dihitung

$<\alpha \rightarrow$ tolak H_0.

Bagian 5

Statistika Nonparametrik

Statistika Parametrik vs Statistika Nonparametrik

- Statistika Parametrik:
  - Teknik-teknik statistika yang didasarkan atas asumsi mengenai populasi yang diambil sampelnya. Contoh: pada uji t diasumsikan populasi terdistribusi normal. Sebutan parametrik digunakan karena pada uji t ini yang diuji adalah parameter (yaitu rata-rata populasi)
  - Membutuhkan data kuantitatif dengan level interval atau rasio

Statistika Parametrik vs Statistika Nonparametrik (lanjutan)

- Statistika Nonparametrik:
  - Cocok untuk data yang tidak memenuhi asumsi statistika parametrik atau yang berjenis kualitatif
  - Disebut juga *distribution-free statistics*
  - Didasarkan atas lebih sedikit asumsi mengenai populasi dan parameter dibandingkan dengan statistika parametrik.
  - Ada yang dapat digunakan untuk data nominal
  - Ada yang dapat digunakan untuk data ordinal

Keuntungan Statistika Nonparametrik

- Kadang-kadang tidak ada alternatifnya pada statistika parametrik
- Uji nonparametrik tertentu dapat digunakan untuk analisis data nominal
- Uji nonparametrik tertentu dapat digunakan untuk analisis data ordinal
- Proses perhitungan pada statistika nonparametrik biasanya lebih sederhana dibandingkan pada statistika parametrik, khususnya untuk sampel kecil

Kekurangan Statistika Nonparametrik

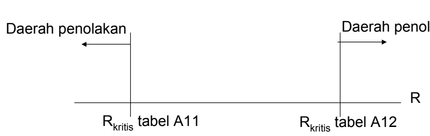
- Uji nonparametrik menjadi tak berguna apabila uji parametrik untuk data yang sama tersedia
- Uji nonparametrik pada umumnya tidak tersedia secara luas dibandingkan dengan uji parametrik
- Untuk sampel besar, perhitungan untuk statistika nonparametrik menjadi rumit

Runs Test

- *Runs Test* satu sampel adalah pengujian nonparametrik untuk menguji keacakan (*randomness*)
- H_0: pengamatan pada sampel terjadi secara acak
- H_a: pengamatan pada sampel terjadi secara tidak acak
- Ide:
  - PWPWPWPWPWPWPWPWPW → tidak acak (banyaknya *runs* = 18)
  - PPPPPPPPPWWWWWWWWW → tidak acak (banyaknya *runs* = 2)
  - Jadi: jika runs terlalu banyak atau terlalu sedikit → tidak acak

Runs Test dengan Sampel Kecil

- Sampel kecil: $n_1 < 20$ dan $n_2 < 20$
- R = banyaknya *runs*
- R_{kritis} pada Tabel A11: $P(R \leq R_{kritis}) < 0.025$
- R_{kritis} pada Tabel A12: $P(R \geq R_{kritis}) < 0.025$
- 0.025 adalah α/2. Jadi α = 0.05.



Contoh

- Apakah *sequence* ini terjadi secara acak? α = 0.05. DCCCCCDCCDCCCCDCDCCCDDDCCC

JAWAB

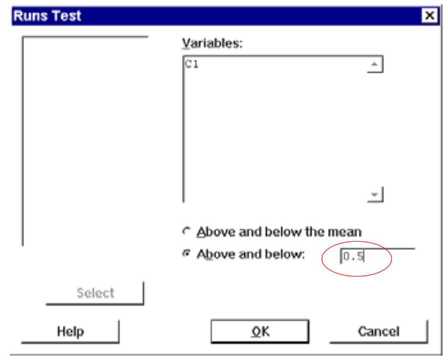
- H_0: pengamatan pada sampel terjadi secara acak
- H_a: pengamatan pada sampel terjadi secara tidak acak
- n_1 = 18 (banyaknya C)
- n_2 = 8 (banyaknya D)
- R = 12
- Dengan n_1 = 18 dan n_2 = 8:
  - dari tabel A11: $R_{kritis} = 7$
  - dari tabel A12: $R_{kritis} = 17$
- Jadi, daerah penolakan adalah $R \leq 7$ dan $R \geq 17$. Karena R = 12 berada di luar daerah penolakan, maka H_0 diterima. Artinya, sequence tersebut terjadi secara acak

Solusi dengan MINITAB

- Dapat digunakan untuk sampel kecil maupun besar
- Ubah data menjadi 1 dan 0 saja, tulis di sebuah kolom

```
Data Display

C1
    1    0    0    0    0    0    1    0    0    1    0    0
    0    0    1    0    1    0    0    0    1    1    1    0
    0    0
```

- Stat → Nonparametrics → Runs Test




```
Runs Test: C1

    C1

    K =      0.5000

    The observed number of runs =   12
    The expected number of runs =   12.0769
     8 Observations above K    18 below
  * N Small -- The following approximation may be invalid
              The test is significant at  0.9710
              Cannot reject at alpha = 0.05
```

sama dengan yang telah diperoleh, R

Ekivalen dengan p-value (nilai p)

Karena p-value > α, maka pertahankan H_o. Artinya urutan data tersebut memang acak

Runs Test dengan Sampel Besar

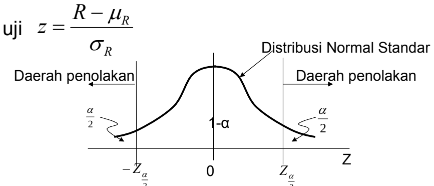
- Untuk n_1 dan n_2 besar, distribusi sampling untuk R akan mendekati distribusi normal dengan rata-rata dan deviasi standar sbb:

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

Runs Test dengan Sampel Besar (lanjutan)

- H_0: pengamatan pada sampel terjadi secara acak
- H_a: pengamatan pada sampel terjadi secara tidak acak
- Statistik uji $z = \dfrac{R-\mu_R}{\sigma_R}$



Daerah penolakan : $|Z| > Z_{\frac{\alpha}{2}}$

Runs Test dengan Sampel Besar (lanjutan)

- Apakah *sequence* ini terjadi secara acak? Gunakan α = 5%
- NNN F NNNNNNN F NN FF NNNNNN F NNNN F NNNNNN FFFF NNNNNNNNNNNN

JAWAB

- H_0: pengamatan pada sampel terjadi secara acak
- H_a: pengamatan pada sampel terjadi secara tidak acak
- n_1 = 40 (banyaknya N)
- n_2 = 10 (banyaknya F)
- R = 13 (banyaknya *runs*)
- Statistik uji $z = \dfrac{R-\mu_R}{\sigma_R}$

$$\mu_R = \frac{2n_1n_2}{n_1+n_2} + 1 = 17$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = 2.213$$

$$z = \frac{13-17}{2.213} = -1.81$$

- Dengan α = 0.05, daerah penolakan adalah jika $|z| > z_{0.025} = 1.96$.
- Karena z = -1.81 berada di luar daerah penolakan, maka pertahankan H_0. Artinya, data tersebut memang terjadi secara acak.
- Dengan MINITAB: Stat → Nonparametrics → Runs Test

```
Runs Test: C1

    C1

    K =      0.5000


    The observed number of runs =  13
    The expected number of runs =  17.0000
    40 Observations above K   10 below
 * N Small -- The following approximation may be invalid
              The test is significant at  0.0707
              Cannot reject at alpha = 0.05
```

Karena p-value > α, maka pertahankan H_0. Artinya urutan data tersebut memang acak

Ekivalen dengan p-value (nilai p)

Mann-Whitney Test (Uji U)

- adalah Uji nonparametrik untuk membandingkan dua populasi independen (pada statistika parametrik: Uji t)
- Populasi tidak harus terdistribusi normal (Pada uji t: harus normal)
- Level data serendah-rendahnya ordinal (uji t tidak dapat)
- Hipotesa yang diuji:
  - H_0: kedua populasi identik
  - H_a: kedua populasi tidak identik

Prosedur Uji U

- Tetapkan satu sampel sebagai Kelompok 1 dan sampel lain sebagai Kelompok 2
- Data dari kedua kelompok disatukan dengan setiap data diberi kode asal kelompoknya
- Data yang telah digabungkan diberi peringkat dari 1 (nilai terkecil) sampai n
- Jumlah peringkat dari kelompok 1 dihitung dan diberi simbol W_1
- Jumlah peringkat dari kelompok 2 dihitung dan diberi simbol W_2
- Langkah selanjutnya: bergantung apakah sampelnya kecil atau besar

Uji U pada Sampel Kecil: $n_1 \leq 10$ dan $n_2 \leq 10$

- Hitung U_1 dan U_2

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - W_1 \quad \text{dan}$$

$$U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

- U adalah yang terkecil di antara U_1 dan U_2. Catatan: salah satu U_i saja yang perlu dihitung, sedangkan U yang satu lagi dapat dihitung dengan $U_j = n_1n_2 - U_i$.
- Gunakan Tabel A13 untuk mendapatkan nilai p untuk U yang telah dihitung. Untuk menggunakan Tabel A13, tetapkan n_1 adalah yang kecil dan n_2 adalah yang besar ($n_1 < n_2$)
- Nilai p pada Tabel A13 adalah untuk uji satu sisi. Untuk uji dua sisi, nilai p nya adalah 2 kali yang ada pada Tabel A13.

Contoh

- Apakah ada perbedaan antara honor per jam pekerja kesehatan dengan pekerja pendidikan? Misalkan diambil sampel acak dari 7 pekerja kesehatan dan 8 pekerja pendidikan. Semua pekerja tersebut diwawancara dan ditanya honor perjamnya, sebagaimana tercantum di dalam tabel berikut. Lakukan pengujian Mann-Whitney U untuk menentukan apakah kedua populasi berbeda di dalam penerimaan honor. Gunakan α = 5%.

Data (sampel)

Pekerja Kesehatan ($)	Pekerja Pendidikan ($)
20.10	26.19
19.80	23.88
22.36	25.50
18.75	21.64
21.90	24.85
22.96	25.30
20.75	24.12
	23.45

Jawab

- Karena populasi tidak dapat diasumsikan normal, maka uji t 2 sampel tidak dapat digunakan (meskipun level data adalah rasio). Jadi digunakan uji U
- H_0: populasi honor pekerja kesehatan dan pekerja pendidikan identik
- H_a: populasi honor pekerja kesehatan dan pekerja pendidikan tidak identik
- n_1 = 7 dan n_2 = 8
- α = 5%

Honor per jam	Peringkat	Kelompok
18.75	1	H
19.80	2	H
20.10	3	H
20.75	4	H
21.64	5	E
21.90	6	H
22.36	7	H
22.96	8	H
23.45	9	E
23.88	10	E
24.12	11	E
24.85	12	E
25.30	13	E
25.50	14	E
26.19	15	E

H = *Health* = Kesehatan, E = *Education* = Pendidikan

$$W_1 = 1+2+3+4+6+7+8 = 31$$

$$W_2 = 5+9+10+11+12+13+14+15 = 89$$

$$U_1 = 7*8 + \frac{7*8}{2} - 31 = 53$$

$$U_2 = 7*8 + \frac{8*9}{2} - 89 = 3$$

atau dihitung dengan 7*8 – 53 = 3

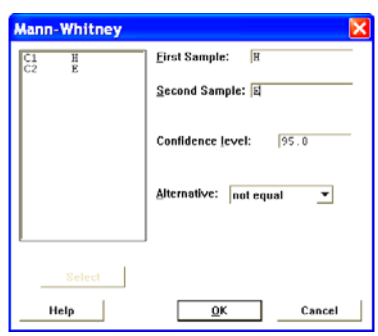
$$U = \min(U_1, U_2) = 3$$

- Dari Tabel A13 untuk n_1 = 7, n_2 = 8, dan U = 3, didapatkan nilai p untuk uji 1 sisi adalah 0.0011. Untuk uji 2 sisi, nilai p = 2 * 0.0011 = 0.0022. Karena nilai p < α, maka tolak H_0. Artinya, populasi honor pekerja kesehatan dan pekerja pendidikan tidak identik. Catatan: terlihat bahwa pada umumnya pekerja pendidikan menerima honor lebih tinggi dari pada pekerja kesehatan

Solusi dengan MINITAB (berlaku untuk sampel kecil maupun besar)

```
Row       H       E

  1   20.10   26.19
  2   19.80   23.88
  3   22.36   25.50
  4   18.75   21.64
  5   21.90   24.85
  6   22.96   25.30
  7   20.75   24.12
  8           23.45
```

• Stat → Nonparametric → Mann-Whitney



```
Mann-Whitney Test and CI: H, E

H          N =   7     Median =      20.750
E          N =   8     Median =      24.485
Point estimate for ETA1-ETA2 is      -3.385
95.7 Percent CI for ETA1-ETA2 is (-5.370,-1.551)
W = 31.0
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is
significant at 0.0046
```

Ekivalen dengan p-value (nilai p)
Sedikit berbeda dengan Tabel A13, hanya karena pembulatan angka

- Karena nilai p < α, maka tolak H_0. Artinya, populasi honor pekerja kesehatan dan pekerja pendidikan tidak identik

Uji U pada Sampel Besar

- Untuk sampel besar ($n_1 > 10$ dan $n_2 > 10$), distribusi sampling untuk U akan mendekati distribusi normal dengan rata-rata dan deviasi standar sebagai berikut:

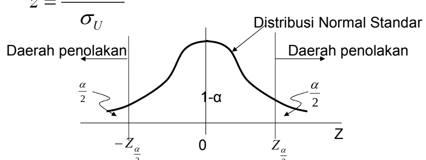
$$\mu_U = \frac{n_1 n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Uji U pada Sampel Besar (lanjutan)

- H_0: kedua populasi identik
- H_a: kedua populasi tidak identik
- Statistik uji

$$z = \frac{U - \mu_U}{\sigma_U}$$



$$\text{Daerah penolakan}: |Z| > Z_{\frac{\alpha}{2}}$$

Contoh

- Apakah uang yang dibelanjakan oleh karyawan untuk makan siang ke restoran sama saja dengan yang ke warung? Untuk menguji hal ini, seorang peneliti mengumpulkan data acak dari karyawan yang makan siang ke restoran dan yang ke warung. Gunakan α = 1%.

Warung ($)	Restoran ($)
2.75	4.10
3.29	4.75
4.53	3.95
3.61	3.50
3.10	4.25
4.29	4.98
2.25	5.75
2.97	4.10
4.01	2.70
3.68	3.65
3.15	5.11
2.97	4.80
4.05	6.25
3.60	3.89
	4.80
	5.50
$n_1 = 14$	$n_2 = 16$

Jawab

- H_0: populasi pengeluaran uang makan siang untuk karyawan yang ke warung sama dengan yang ke restoran
- H_a: populasi pengeluaran uang makan siang untuk karyawan yang ke warung tidak sama dengan yang ke restoran
- $n_1 > 10$ dan $n_2 > 10$, maka gunakan Uji U untuk sampel besar
- α = 0.01. Apabila nilai p < α maka tolak H_0.

Nilai	Peringkat	Kelompok
2.25	1	W
2.70	2	R
2.75	3	W
2.97	4.5	W
2.97	4.5	W
3.10	6	W
3.15	7	W
3.29	8	W
3.50	9	R
3.60	10	W
3.61	11	W
3.65	12	R
3.68	13	W
3.89	14	R
3.95	15	R

Nilai	Peringkat	Kelompok
4.01	16	W
4.05	17	W
4.10	18.5	R
4.10	18.5	R
4.25	20	R
4.29	21	W
4.53	22	W
4.75	23	R
4.80	24.5	R
4.80	24.5	R
4.98	26	R
5.11	27	R
5.50	28	R
5.75	29	R
6.25	30	R

- Jumlah peringkat yang dari kelompok W (Warung) = W_1 = 1+3+4.5+4.5+6+7+8+10+11+13+16+17+21+22 = 144
- $U_1 = 14*16 + \frac{14*15}{2} - 144 = 185$

  $U_2 = 14*16 - 185 = 39$

  $\mathrm{U} = \min(39,185) = 39$

  $\mu_U = \frac{14*16}{2} = 112$

  $\sigma_U = \sqrt{\frac{14*16*31}{12}} = 24.1$

  $z = \frac{39-112}{24.1} = \text{-}3.03$

uji 2 sisi

- Nilai p untuk z = -3.03 adalah 2 * 0.0012 = 0.0024 < α → tolak H_0. Artinya: populasi pengeluaran uang makan siang untuk karyawan yang ke warung tidak sama dengan yang ke restoran
- Dengan MINITAB:

```
Mann-Whitney Test and CI: W, R

W          N =  14     Median =       3.445
R          N =  16     Median =       4.500
Point estimate for ETA1-ETA2 is      -1.065
95.2 Percent CI for ETA1-ETA2 is (-1.700,-0.460)
W = 144.0
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is significant
at 0.0026
The test is significant at 0.0026 (adjusted for ties)
```

p-value

Uji Peringkat Bertanda (Wilcoxon) untuk data Sepadan

- Data Sepadan (*matched pairs*):
  - Statistika Parametrik: Uji t (asumsi: populasi normal)
  - Statistika Nonparametrik: Uji Wilcoxon
- Uji Wilcoxon (seperti juga uji t) digunakan untuk menganalisis data pada 2 kelompok yang berkaitan, termasuk kasus *before-and-after* di mana orang atau objek yang sama diamati pada dua kondisi yang berbeda
- Jenis data pada Wilcoxon: serendah-rendahnya level ordinal
- Asumsi Uji Wilcoxon
  - Pasangan data diambil secara acak
  - Distribusi populasi: simetris

Prosedur Uji Wilcoxon

- n = banyaknya pasangan data
- Urutkan perbedaan antara kedua data (d), dari yang terkecil sampai yang terbesar, tanpa memperhatikan apakah perbedaan tersebut (-) atau (+)
- Jika perbedaan tersebut (-) maka peringkatnya juga diberi tanda (-)
- Perbedaan (d) yang bernilai 0 (apabila ada) diabaikan, dan banyak data (n) dikurangi sebanyak d yang bernilai 0
- Jumlahkan peringkat yang bertanda (-), sebut T-. Tanda (-) tidak ikut didalam perjumlahan
- Jumlahkan peringkat yang bertanda (+), sebut T+.
- Statistik uji: T = min (T- dan T+)

Hipotesa yang diuji pada Uji Wilcoxon

- H_0: $M_d = 0$ versus H_a: $M_d \neq 0$ (*two-tailed test*)
- H_0: $M_d = 0$ versus H_a: $M_d > 0$ (*one-tailed test*)
- H_0: $M_d = 0$ versus H_a: $M_d < 0$ (*one-tailed test*)
- Catatan:
  - M_d = median perbedaan antara kedua populasi
  - M_d = 0 berarti kedua populasi identik

Uji Wilcoxon untuk Sampel Kecil ($n \leq 15$)

- Dengan n dan α, gunakan Tabel A14 (tersedia untuk *one-tailed test* dan *two-tailed test*) untuk mendapatkan T_{kritis}.
- Jika $T \leq T_{kritis}$ → tolak H_0.

Contoh

- Seorang peneliti melakukan survey mengenai biaya pemeliharaan kesehatan yang dikeluarkan oleh keluarga di kota A dan B. Peneliti tersebut mengambil enam pasang keluarga yang dipadankan secara demografis di kota A dan B. Dari keenam pasang keluarga tersebut dicatat biaya pemeliharaan kesehatan pada tahun yang lalu (dalam USD). Dengan menggunakan α = 0.05, lakukan pengujian untuk menentukan apakah ada perbedaan signifikan di dalam pengeluaran biaya kesehatan di antara kedua kota tersebut

Pasangan keluarga	A	B
1	1950	1760
2	1840	1870
3	2015	1810
4	1580	1660
5	1790	1340
6	1925	1765

Jawab

- Karena populasi tidak dapat diasumsikan normal, maka digunakan Uji Wilcoxon (bukan uji t), meskipun datanya berlevel rasio
- H_0: $M_d = 0$ versus H_a: $M_d \neq 0$
- α = 0.05.
- n = 6 ($\leq$ 15) → sampel kecil

Kel	A	B	Perbe-daan d	Pering-kat
1	1950	1760	+190	+4
2	1840	1870	-30	-1
3	2015	1810	+205	+5
4	1580	1660	-80	-2
5	1790	1340	+450	+6
6	1925	1765	+160	+3

- T+ = 4+5+6+3 = 18
- T- = 1+2 = 3
- T = min (T- dan T+) = min (18 dan 3) = 3
- n = 6, α = 0.05 → (Tabel A14, *two-tailed test*) T_{kritis} = 1. Karena $T > T_{kritis}$ maka pertahankan H_0. Artinya tidak cukup bukti bahwa pengeluaran biaya kesehatan di kedua kota berbeda

Uji Wilcoxon untuk Sampel Besar ($n > 15$)

- Untuk sampel besar distribusi sampling untuk T akan mendekati distribusi normal dengan rata-rata dan deviasi standar sebagai berikut:

$$\mu_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

- Statistik uji: $z = \dfrac{T - \mu_T}{\sigma_T}$

Contoh

- Sebuah perusahaan berupaya meningkatkan produktivitas dengan menerapkan kontrol kualitas. Untuk meneliti apakah penerapan kontrol kualitas tersebut memang berhasil meningkatkan produksi, diambil sampel dari 20 pekerja dan dicatat produksi dari masing-masing pekerja sebelum dan sesudah penerapan kontrol kualitas tersebut. Gunakan Uji Wilcoxon dan α = 0.01 untuk membuktikan apakah kontrol kualitas tersebut memang berhasil meningkatkan produksi.

Pekerja	Before	After	d = Before – After	Peringkat
1	5	11	-6	-19
2	4	9	-5	-17
3	9	9	0	Hapus
4	6	8	-2	-9
5	3	5	-2	-9
6	8	7	1	+3.5
7	7	9	-2	-9
8	10	9	1	+3.5
9	3	7	-4	-14.5
10	7	9	-2	-9
11	2	6	-4	-14.5
12	5	10	-5	-17
13	4	9	-5	-17
14	5	7	-2	-9
15	8	9	-1	-3.5
16	7	6	1	+3.5
17	9	10	-1	-3.5
18	5	8	-3	-12.5
19	4	5	-1	-3.5
20	3	6	-3	-12.5

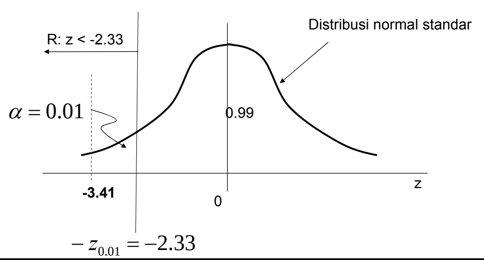
- H_0: $M_d = 0$ versus H_a: $M_d < 0$
- T- = 179.5
- T+ = 10.5
- T = min(179.5, 10.5) = 10.5
- n = 19 (1 data dengan d = 0 dihapus)
- Menghitung statistik uji:

$$\mu_T = \frac{n(n+1)}{4} = \frac{19*20}{4} = 95$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{19*20*39}{24}} = 24.8$$

$$z = \frac{T-\mu_T}{\sigma_T} = \frac{10.5-95}{24.8} = -3.41$$

- Dengan α = 0.01, daerah penolakan: $z < -z_{0.01} = -2.33$
- Karena z terletak di daerah penolakan (-3.41 < -2.33), maka tolak H_0. Artinya: memang benar bahwa setelah ada program kontrol kualitas, produktivitas meningkat
- Dengan MINITAB: Stat → Nonparametric → 1 sample Wilcoxon



Row	Before	After	d
1	5	11	-6
2	4	9	-5
3	9	9	0
4	6	8	-2
5	3	5	-2
6	8	7	1
7	7	9	-2
8	10	9	1
9	3	7	-4
10	7	9	-2
11	2	6	-4
12	5	10	-5
13	4	9	-5
14	5	7	-2
15	8	9	-1
16	7	6	1
17	9	10	-1
18	5	8	-3
19	4	5	-1
20	3	6	-3


```
Wilcoxon Signed Rank Test: d

Test of median = 0.000000 versus median  <  0.000000

              N for   Wilcoxon              Estimated
          N    Test  Statistic        P        Median
d        20      19       10.5    0.000        -2.000
```

Statistik Uji: T

p-value. Karena p-value < α, maka tolak H_0.

Uji Kruskal-Wallis

- Statistika Parametrik: Anova Satu Arah. Asumsi:
  - Populasi terdistribusi normal
  - Setiap kelompok Independen
  - Varians populasi sama
  - Data acak
- Statistika Nonparametrik: Uji Kruskal-Wallis. Asumsi:
  - Tidak ada asumsi tentang bentuk populasi
  - Setiap kelompok Independen
  - Data acak
- Uji Kruskal-Wallis: menentukan apakah semua kelompok berasal dari populasi yang sama, ataukah sedikitnya satu kelompok berasal dari populasi yang berbeda
- Banyak kelompok = c (>2)

Prosedur Uji Kruskal-Wallis

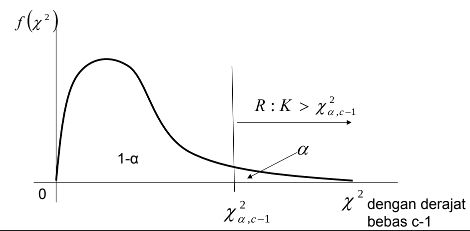
- Data dari setiap kelompok diberi peringkat dari 1 (terkecil), dengan memandang seolah-olah semuanya berasal dari 1 kelompok.
- Hitung statistik uji K:

$$K = \frac{12}{n(n+1)}\left(\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right) - 3(n+1)$$

  - c = banyaknya kelompok
  - n = total banyaknya *items*
  - T_j = total peringkat pada satu kelompok j
  - n_j = banyaknya *items* pada satu kelompok j
  - K terdistribusi χ^2 dengan df = c-1

Prosedur Uji Kruskal-Wallis (lanjutan)

- H_0: seluruh c populasi identik
- H_a: sedikitnya 1 populasi berbeda
- Daerah penolakan: selalu di kanan, yaitu: R: K > $\chi^2_{\alpha,\, c-1}$



Contoh

- Seorang peneliti dalam bidang agrobisnis tertarik untuk menentukan kondisi yang dapat menyebabkan pertumbuhan bibit cemara secara lebih cepat. Ia mencoba pada 24 bibit cemara yang diberi kondisi berbeda (lihat tabel). Hasil pengamatan setelah setahun adalah tinggi bibit (dalam in.). Dengan menggunakan α = 0.01, lakukan Uji Kruskal-Wallis untuk menentukan apakah ada perbedaan signifikan pada keempat kondisi tersebut terhadap pertumbuhan bibit cemara.

Data pengamatan

Kelompok 1: alami	Kelompok 2: ditambah air	Kelompok 3: ditambah vertilizer	Kelompok 4: ditambah air & vertilizer
8	10	11	18
5	12	14	20
7	11	10	16
11	9	16	15
9	13	17	14
6	12	12	22

Peringkat

K1	K2	K3	K4
4	7.5	10	22
1	13	16.5	23
3	10	7.5	19.5
10	5.5	19.5	18
5.5	15	21	16.5
2	13	13	24

$$T_1 = 25.5 \quad T_2 = 64.0 \quad T_3 = 87.5 \quad T_4 = 123$$

$$n_1 = 6 \quad n_2 = 6 \quad n_3 = 6 \quad n_4 = 6$$

$$\sum_{j=1}^{4}\frac{T_j^2}{n_j} = \frac{25.5^2}{6} + \frac{64^2}{6} + \frac{87.5^2}{6} + \frac{123^2}{6} = 4588.6$$

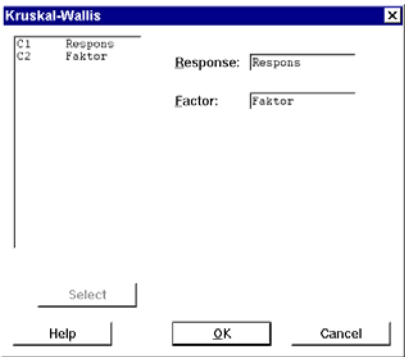
$$K = \frac{12}{24*25}(4588.6) - 3(24+1) = 16.77$$

- df = 4 – 1 = 3. α = 0.01. Daerah penolakan R: K > $\chi^2_{0.01,3}$ = 11.345.
- Karena K ada di R, maka tolak H_0. Artinya ada perbedaan signifikan pada berbagai kondisi terhadap pertumbuhan bibit cemara

Dengan MINITAB

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

Stat → Nonparametric → Kruskal-Wallis




```
Kruskal-Wallis Test: Respons versus Faktor

Kruskal-Wallis Test on Respons

Faktor      N    Median    Ave Rank        Z
1           6     7.500         4.3    -3.30
2           6    11.500        10.7    -0.73
3           6    13.000        14.6     0.83
4           6    17.000        20.5     3.20
Overall    24                  12.5

H = 16.77  DF = 3  P = 0.001
H = 16.86  DF = 3  P = 0.001 (adjusted for ties)
```

statistik uji: K

p-value. Karena p-value <α, maka tolak H_0.

Uji Friedman

- Statistika Parametrik: *randomized block design*. Asumsi: populasi terdistribusi normal, data interval atau rasio
- Statistika Nonparametrik: uji Friedman. Asumsi: populasi tidak harus terdistribusi normal, data serendah-rendahnya peringkat
- Asumsi lain pada Uji Friedman:
  - Setiap blok independen
  - Tidak ada interaksi antara blok dan *treatment*
  - Pengamatan di dalam setiap blok dapat dijadikan peringkat

Prosedur Uji Friedman

- H_0: Populasi *treatment* sama
- H_a: Sedikitnya satu populasi *treatment* menghasilkan nilai lebih besar dari sedikitnya satu populasi *treatment* lain
- Hitung peringkat di dalam setiap blok (tidak dicampur dengan blok lain), kecuali apabila datanya memang berlevel peringkat

Statistik Uji pada Uji Friedman

$$\chi^2 = \frac{12}{bc(c+1)} \sum_{j=1}^{c} R_j^2 - 3b(c+1)$$

- df = c - 1
- c = banyaknya kolom (*treatment levels*)
- b = banyaknya baris (blok)
- R_j = total peringkat pada kolom j; j = 1, 2, … c

Contoh

- Sebuah riset pemasaran ingin mempelajari kinerja lemari es dari 5 merk yang berbeda (merk A, B, C, D, dan E). Untuk itu, sepuluh orang yang berpotensi menjadi pembeli lemari es diminta memberi peringkat pada kelima merk lemari es tersebut. Gunakan Uji Friedman dan α = 0.01 untuk menentukan apakah ada perbedaan yang signifikan pada peringkat kelima merk lemari es tersebut.

Orang	Merk 1	Merk 2	Merk 3	Merk 4	Merk 5
1	3	5	2	4	1
2	1	3	2	4	5
3	3	4	5	2	1
4	2	3	1	4	5
5	5	4	2	1	3
6	1	5	3	4	2
7	4	1	3	2	5
8	2	3	4	5	1
9	2	4	5	3	1
10	3	5	4	2	1
R_j	26	37	31	31	25
R_j^2	676	1369	961	961	625

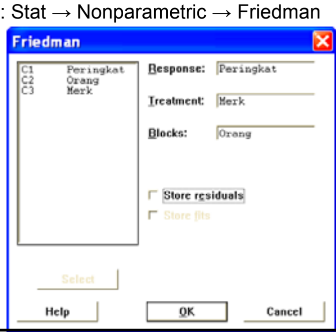
$$\sum_{j=1}^{5} R_j^2 = 4592$$

Jawab

- H_0: Populasi kelima merk sama
- H_a: Sedikitnya satu populasi merk berperingkat lebih tinggi dibandingkan populasi merk lainnya
- b = 10
- c = 5
- df = c – 1 = 5 – 1 = 4
- α = 0.01
- Dengan α = 0.01 dan df = 4, didapatkan $\chi^2_{0.01,4}$ = 13.2767. Jadi tolak H_0 apabila $\chi^2 > 13.2767$.

$$\chi^2 = \frac{12}{bc(c+1)}\sum_{j=1}^{c} R_j^2 - 3b(c+1) = \frac{12}{10*5*6}4592 - 3*10*6 = 3.68$$

- Karena $\chi^2 < 13.2767$, maka pertahankan H_0 Artinya, dari kelima merk tersebut, tidak ada yang kinerjanya menonjol dibandingkan lainnya
- MINITAB: Stat → Nonparametric → Friedman




```
Friedman Test: Peringkat versus Merk, Orang

Friedman test for Peringka by Merk blocked by
Orang

S = 3.68  DF = 4  P = 0.451

                   Est   Sum of
Merk          N  Median   Ranks
1            10   2.300    26.0
2            10   4.000    37.0
3            10   3.000    31.0
4            10   3.000    31.0
5            10   1.700    25.0

Grand median  =   2.800
```

statistik uji χ^2

p-value.
Karena p-value >α,
maka pertahankan H_0.

Korelasi Peringkat Spearman

- Ukuran asosiasi antara dua variabel yang berjenis interval atau rasio: koefisien korelasi Person
- Untuk dua variabel berjenis ordinal, ukuran asosiasinya adalah koefisien korelasi Spearman

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

- n = banyaknya pasangan data yang dicari korelasinya
- d = perbedaan peringkat pada setiap pasang. Di setiap kelompok dibuat peringkatnya dari 1 sampai n.
- Interpretasi r_s sama saja dengan interpretasi r

Contoh

- Apakah ada hubungan kuat antara harga minyak mentah (per barrel) dan harga BBM (per galon) di pompa bensin? Untuk mengestimasi asosiasi antara kedua variabel tersebut, seorang peneliti di perusahaan minyak mengunpulkan data di sebuah kota selama 9 bulan, dan mencatat rata-rata harga di setiap bulan tersebut. Hitunglah koefisien korelasi Spearman untuk data ini.

Row	Mentah	BBM	Mentah_P	BBM_P	d	d2
1	14.60	1.05	3	1.0	2.0	4.00
2	10.58	1.06	1	2.5	-1.5	2.25
3	12.30	1.08	2	4.0	-2.0	4.00
4	15.10	1.06	4	2.5	1.5	2.25
5	18.35	1.12	5	5.0	0.0	0.00
6	22.60	1.24	6	6.0	0.0	0.00
7	28.90	1.36	8	8.0	0.0	0.00
8	31.40	1.40	9	9.0	0.0	0.00
9	26.75	1.34	7	7.0	0.0	0.00

hasil pengamatan **peringkat** **perbedaan peringkat**

$$\sum d^2 = 12.5$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6*12.5}{9(9^2-1)} = 0.89583$$

Solusi dengan MINITAB

- Tulis data di ‘Mentah’ dan ‘BBM’
- Data → Rank. Rank data in ‘Mentah’, Store ranks in ‘Mentah_P’
- Data → Rank. Rank data in ‘BBM’, Store ranks in ‘BBM_P’
- Stat → Basic Statistics → Correlation. Variables: ‘Mentah_P’ ‘BBM_P’

Correlations: Mentah_P, BBM_P

Pearson correlation of Mentah_P and BBM_P = 0.895
P-Value = 0.001

Bagian 6

Peramalan dengan Deret Waktu (Time Series)

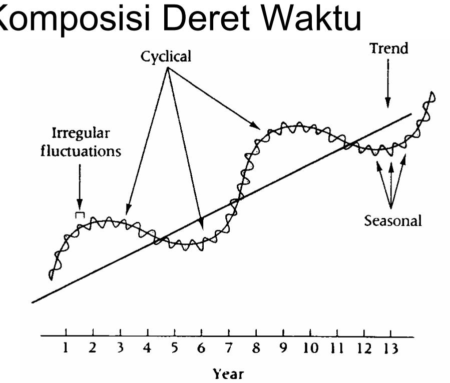
Peramalan (*Forecasting*)

- adalah seni dan pengetahuan untuk memrediksi masa depan. Peramalan digunakan di dalam proses pengambilan keputusan untuk membantu pebisnis menyimpulkan tentang pembelian, penjualan, produksi, dll. Contoh:
  - Pengamat pasar memprediksi nilai saham di tahun depan
  - Perencana kota meramalkan krisis air di suatu kota
  - Harga BBM akan meningkat secara tajam pada beberapa bulan yad

Data Deret Waktu

- adalah data yang dikumpulkan mengenai suatu karakteristik tertentu pada suatu periode waktu atau interval yang teratur
- digunakan untuk memrediksi sesuatu di masa yang akan datang

Komposisi Deret Waktu



Komposisi Deret Waktu

- *Trend*: arah umum jangka panjang suatu data
- *Cycle*: pola tinggi rendahnya data pada periode waktu yang lebih dari satu tahun
- *Seasonal effects*: siklus data yang terjadi pada periode waktu kurang dari 1 tahun
- *Irregular fluctuations*: perubahan cepat pada data pada selang waktu jauh lebih pendek dibandingkan *seasonal effects*

Pengukuran Galat pada Peramalan

- Galat peramalan individual:

$$e_t = x_t - F_t$$

e_t = galat pada peramalam
x_t = nilai aktual
F_t = nilai peramalan

- Deviasi Mutlak Rata-rata (*Mean Absolute Deviation = MAD)*:

$$MAD = \frac{\sum |e_i|}{banyaknya\ peramalan}$$

Pengukuran Galat pada Peramalan (lanjutan)

- Galat Kuadrat Rata-rata (*Mean Square Error = MSE*):

$$MSE = \frac{\sum e_i^2}{banyaknya\ peramalan}$$

- Pemilihan pengukuran galat pada peramalan bergantung pada peneliti. Masing-masing cara menghasilkan informasi yang berbeda.

Contoh perhitungan MAD dan MSE

i	Aktual	Peramalan	e_i	abs(e)	e^2
1	19.4	16.6	2.8	2.8	7.8
2	23.6	19.1	4.5	4.5	20.3
3	24.0	22.0	2.0	2.0	4.0
4	26.8	24.8	2.0	2.0	4.0
5	29.2	25.9	3.3	3.3	10.9
6	35.5	28.6	6.9	6.9	47.6
			Jumlah	21.5	94.6

$$MAD = \frac{21.5}{6} = 3.6 \qquad MSE = \frac{94.6}{6} = 15.8$$

Cara-cara Penghalusan (*Smoothing Techniques*)

- adalah cara-cara untuk menghilangkan efek tak teratur pada data deret waktu.
- antara lain:
  - Model peramalan naif
  - Model Perataan
  - Penghalusan eksponensial

Model peramalan naif

- Adalah model sederhana yang menggunakan asumsi bahwa data pada periode waktu yang lebih mutakhir merepresentasikan prediksi atau peramalan untuk masa yang akan datang.
- Cocok untuk data deret waktu yang selang waktunya adalah harian atau mingguan, atau yang tidak menunjukkan *trend* atau *seasonality.*

$$F_t = x_{t-1}$$

F_t = nilai peramalan untuk periode waktu t
x_{t-1} = nilai untuk periode waktu t-1

Model Perataan

- Dihitung dengan menggunakan rata-rata dari beberapa periode waktu dan menggunakan rata-rata sebagai peramalan untuk periode waktu berikutnya
- Contoh:
  - Rata-rata Sederhana
  - Rata-rata Bergerak
  - Rata-rata Bergerak Berbobot

Rata-rata Sederhana (*Simple Average*)

- Peramalan untuk periode waktu t adalah rata-rata dari nilai sejumlah tertentu periode waktu di masa lalu:

$$F_t = \frac{X_{t-1} + X_{t-2} + X_{t-3} + \ldots\ldots + X_{t-n}}{n}$$

Rata-rata Bergerak (*Moving Average*)

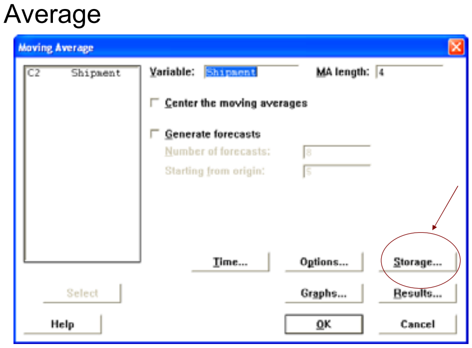
- Adalah rata-rata yang diperbarui atau dihitung ulang untuk setiap periode waktu yang baru yang ditinjau.
- Keuntungan: Informasi yang lebih baru digunakan pada setiap rata-rata bergerak yang baru.
- Kerugian:
  - Sulit untuk menentukan panjang waktu yang optimal untuk menghitung rata-rata bergerak
  - Rata-rata bergerak biasanya tidak mengoreksi efek-efek deret waktu seperti *trend, cycles*, dan *seasonality*.
- Untuk menentukan waktu yang optimal: gunakan panjang waktu yang berbeda-beda, lalu bandingkan galatnya.

Contoh Rata-rata Bergerak 4 bulan

Month	*Shipment*	*Average*	*Error*
Jan	1056		
Feb	1345		
Mar	1381		
Apr	1191		
May	1259	1243.25	15.75
Jun	1361	1294.00	67.00
Jul	1110	1298.00	-188.00
Aug	1334	1230.25	103.75
Sep	1416	1266.00	150.00
Oct	1282	1305.25	-23.25
Nov	1341	1285.50	55.50
Dec	1382	1343.25	38.75

output

- MINITAB: Stat -> Time Series -> Moving Average



Moving Average for Shipment

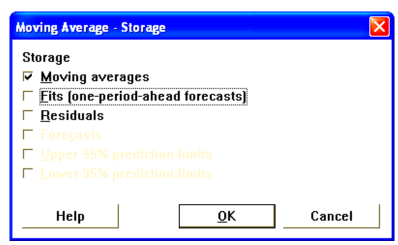
Data Shipment
Length 12
NMissing 0

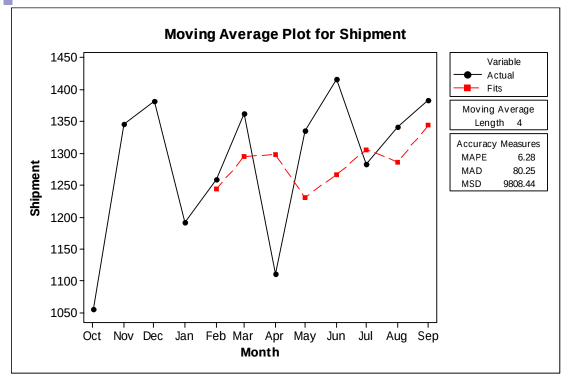
Moving Average

Length 4

Accuracy Measures

MAPE 6.28
MAD 80.25
MSD 9808.44





Rata-rata Bergerak Berbobot (*Weighted Moving Average*)

- Adalah rata-rata bergerak yang menggunakan bobot yang berbeda antara suatu periode waktu dengan periode waktu lainnya.
- Pembagi (penyebut) adalah jumlah total bobot untuk setiap periode waktu.
- Contoh: misalnya untuk rata-rata bergerak berbobot 3 bulan, bobot untuk bulan ke 1 adalah 1, ke 2 adalah 2, dan ke tiga, adalah 3. Rumusnya adalah:

$$\bar{x}_{berbobot} = \frac{3M_{t-1} + 2M_{t-2} + M_{t-3}}{6}$$

Contoh Rata-rata Bergerak Berbobot

- Untuk data *shipment* di atas, carilah rata-rata bergerak berbobot dengan menggunakan bobot: 4 untuk bulan terakhir, 2 untuk bulan sebelumnya, dan 1 untuk bulan lainnya. Panjang waktu untuk rata-rata bergerak adalah 4 bulan.
- Rumus umum untuk kasus ini:

$$\bar{x}_{berbobot} = \frac{4M_{t-1} + 2M_{t-2} + M_{t-3} + M_{t-4}}{8}$$

Contoh Rata-rata Bergerak Berbobot (lanjutan)

Month	*Shipment*	*Average*	*Error*
Jan	1056		
Feb	1345		
Mar	1381		
Apr	1191		
May	1259	1240.88	18.13
Jun	1361	1268.00	93.00
Jul	1110	1316.75	-206.75
Aug	1334	1201.50	132.50
Sep	1416	1272.00	144.00
Oct	1282	1350.38	-68.38
Nov	1341	1300.50	40.50
Dec	1382	1334.75	47.25

Penghalusan Eksponensial

- Digunakan untuk membobotkan data dari periode-periode waktu sebelumnya, dengan taraf kepentingan yang berkurang secara eksponensial di dalam peramalan.
- Dilakukan dengan mengalikan nilai aktual dengan konstanta penghalusan eksponensial di antara 0 dan 1 yang diberi simbol α.

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t$$

  - F_{t+1} = peramalan untuk periode waktu berikutnya (t+1)
  - F_t = peramalan untuk periode waktu saat ini (t)
  - X_t = nilai aktual untuk periode waktu saat ini
  - α = nilai antara 0 dan 1 yang disebut dengan konstanta penghalusan eksponensial

Contoh Penghalusan Eksponensial

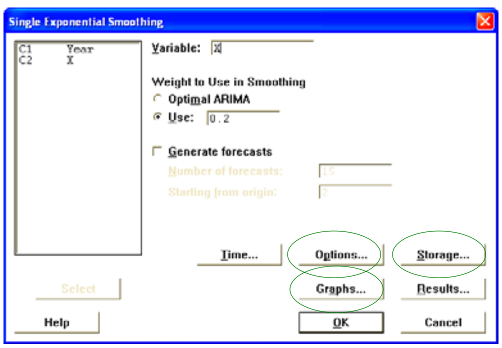
- Untuk data tahunan X berikut ini (dari 1984 sampai dengan 1999), gunakanlah penghalusan eksponensial untuk meramalkan nilai untuk setiap periode waktu. Gunakanlah α = 0.2, 0.5, dan 0.8

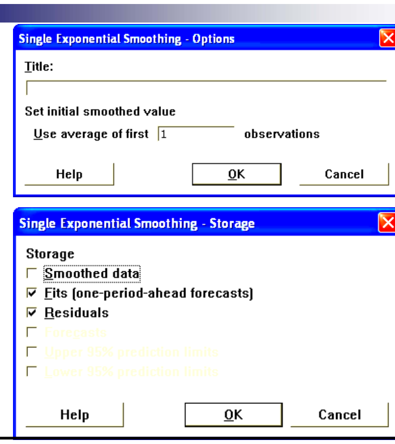
Year	X	α = 0.2		α = 0.5		α = 0.8	
		F	e	F	e	F	e
1984	1750	-	-	-	-	-	-
1985	1742	1750.0	-8.0	1750.0	-8.0	1750.0	-8.0
1986	1805	1748.4	56.6	1746.0	59.0	1743.6	61.4
1987	1620	1759.7	-139.7	1775.5	-155.5	1792.7	-172.7
1988	1488	1731.8	-243.8	1697.8	-209.8	1654.5	-166.5
1989	1376	1683.0	-307.0	1592.9	-216.9	1521.3	-145.3
1990	1193	1621.6	-428.6	1484.4	-291.4	1405.1	-212.1
1991	1014	1535.9	-521.9	1338.7	-324.7	1235.4	-221.4
1992	1200	1431.5	-231.5	1176.4	23.6	1058.3	141.7
1993	1288	1385.2	-97.2	1188.2	99.8	1171.7	116.3
1994	1457	1365.8	91.2	1238.1	218.9	1264.7	192.3
1995	1354	1384.0	-30.0	1347.5	6.5	1418.5	-64.5
1996	1477	1378.0	99.0	1350.8	126.2	1366.9	110.1
1997	1474	1397.8	76.2	1413.9	60.1	1455.0	19.0
1998	1617	1413.0	204.0	1443.9	173.1	1470.2	146.8
1999	1666	1453.8	212.2	1530.5	135.5	1587.6	78.4

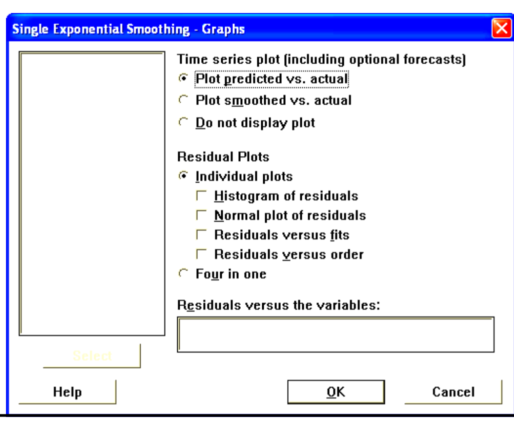
Contoh perhitungan untuk $\alpha = 0.2$

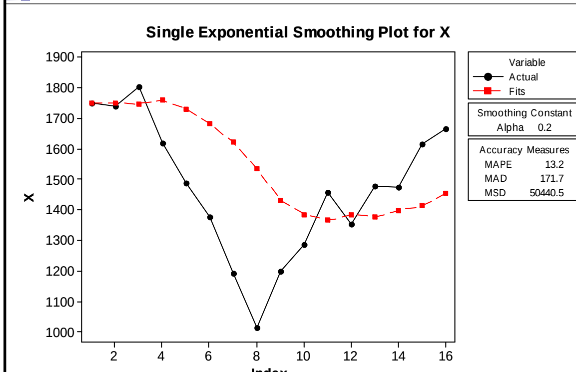
- 1984: F belum ada
- 1985: F = mengambil data aktual tahun 1984
- 1986: $F = 0.2X_{1985} + 0.8F_{1985} = 0.2*1742 + 0.8*1750 = 1748.4$
- 1987: $F = 0.2X_{1986} + 0.8F_{1986} = 0.2*1805 + 0.8*1748.4 = 1759.7$
- e = X – F setiap tahun

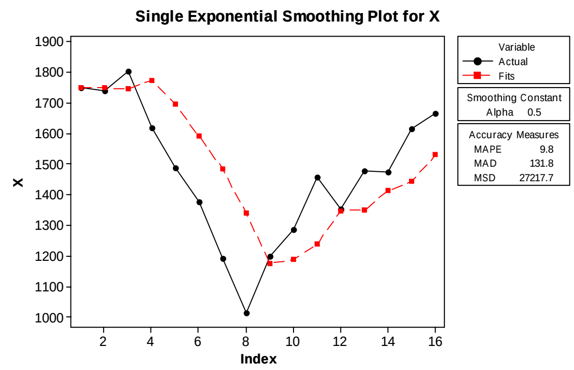
- MINITAB: Stat -> Time Series -> Single Exp. Smoothing

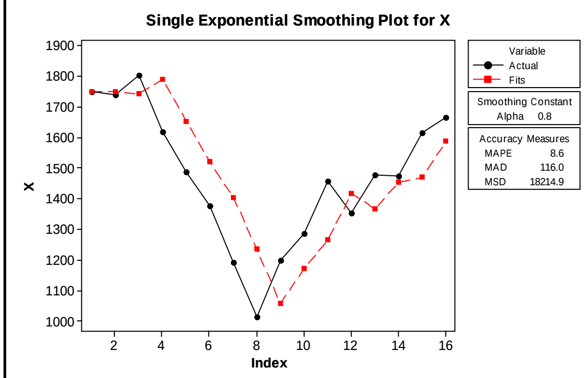












Analisis *Trend*

- *Trend* adalah arah umum jangka panjang dari suatu besaran pada suatu periode yang lebih dari 1 tahun (biasanya beberapa tahun).
- Salah satu cara analisis *trend* adalah dengan analisis regresi, dengan:
  - Y = besaran yang diramalkan
  - X = periode waktu

    Catatan: Misalkan data yang ada adalah untuk tahun 1981 sampai 2000. Maka X adalah 1 sampai 20, bukan 1981 sampai 2000.
- Di dalam analisis *trend*, efek musim *(seasonal effects)* diasumsikan tidak ada, atau sudah dieliminasi.

Efek Musim (*Seasonal Effects*)

- Efek musim adalah pola perilaku data yang terjadi pada periode waktu kurang dari 1 tahun.
- Dekomposisi dengan model perkalian:

  T*C*S*I
  - T = trend
  - C = cyclicality
  - S = seasonality
  - I = irregularity

Langkah dekomposisi

- Hilangkan efek T dan C dari setiap data sehingga:

$$\frac{T*C*S*I}{T*C} = S*I$$

- Hilangkan efek I sehingga hanya tersisa efek S

$$S = \frac{S*I}{I}$$

Quarter	Actual Values $(T \cdot C \cdot S \cdot I)$	4-Quarter Moving Total	4-Quarter 2-Year Moving Total	Ratios of Actual Centered Moving Average $(T \cdot C)$	Values to Moving Averages $(S \cdot I) \cdot (100)$
1 (year 1)	4,009				
2	4,321				
3	4,224	16,498	33,110	4,139	102.05
4	3,944	16,612	33,425	4,178	94.40
1 (year 2)	4,123	16,813	34,059	4,257	96.85
2	4,522	17,246	34,578	4,322	104.63
3	4,657	17,332	35,034	4,379	106.35
4	4,030	17,702	35,688	4,461	90.34
1 (year 3)	4,493	17,986	35,866	4,483	100.22
2	4,806	17,880	36,215	4,527	106.16
3	4,551	18,335	36,772	4,597	99.00
4	4,485	18,437	36,867	4,608	97.33
1 (year 4)	4,595	18,430	36,726	4,591	100.09
2	4,799	18,296	36,365	4,546	105.57
3	4,417	18,069	35,788	4,474	98.73
4	4,258	17,719	35,539	4,442	95.86
1 (year 5)	4,245	17,820	35,808	4,476	94.84
2	4,900	17,988	36,251	4,531	108.14
3	4,585	18,263			
4	4,533				

dibagi 8

TCSI/ TC *100

Indeks Musim

Quarter	Thn 1	Thn 2	Thn 3	Thn 4	Thn 5
1	-	96.85	100.22	100.09	94.84
2	-	104.63	106.16	105.57	108.14
3	102.05	106.35	99.00	98.71	-
4	94.40	90.34	97.33	95.86	-

Quarter	Index
1	98.47
2	105.87
3	100.53
4	95.13

$$\frac{96.85+100.09}{2}$$

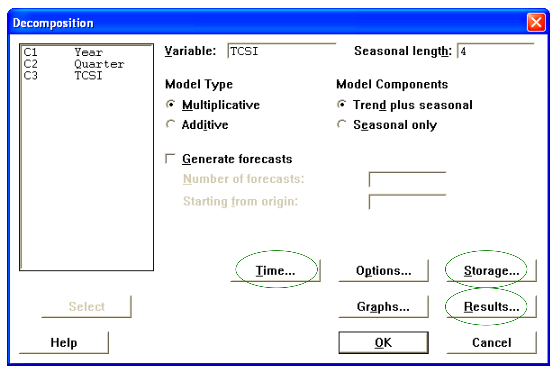
Tidak ikut dirata-rata (yang terbesar dan terkecil)

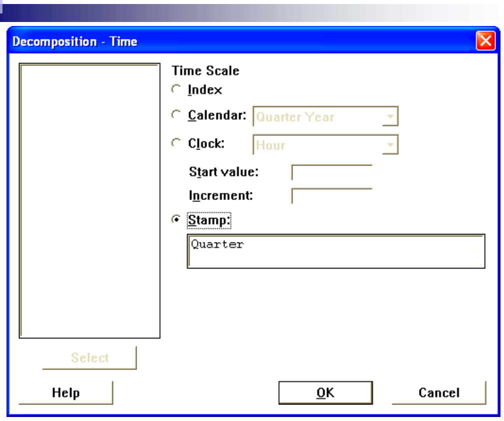
Tahun	Quarter	Nilai Aktual (T*C*S*I)	Indeks Musim	Data tanpa efek musim (*Deseasonalized data*) (T*C*I)
1	1	4009	98.47	4071
	2	4321	105.87	4081
	3	4224	100.53	4202
	4	3944	95.13	4146
2	1	4123	98.47	4187
	2	4522	105.87	4271
	3	4657	100.53	4632
	4	4030	95.13	4236
3	1	4493	98.47	4563
	2	4806	105.87	4540
	3	4551	100.53	4327
	4	4485	95.13	4715

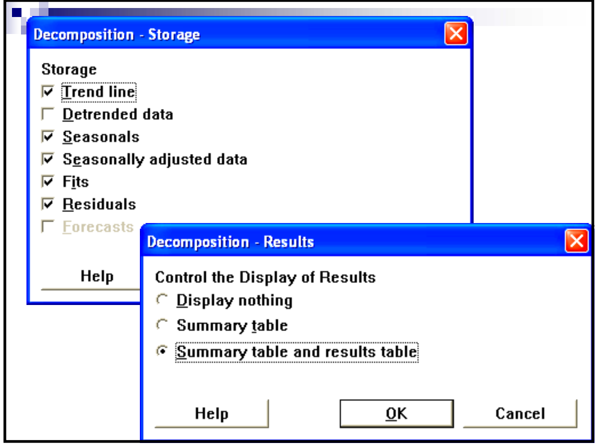
bersambung

Tahun	Quarter	Nilai Aktual (T*C*S*I)	Indeks Musim	Data tanpa efek musim (*Deseasonalized data*) (T*C*I)
4	1	4595	98.47	4666
	2	4799	105.87	4533
	3	4417	100.53	4394
	4	4258	95.13	4476
5	1	4245	98.47	4311
	2	4900	105.87	4628
	3	4585	100.53	4561
	4	4533	95.13	4765

- MINITAB: Stat -> Time Series -> Decomposition







Time Series Decomposition for TCSI

Multiplicative Model
Data   TCSI
Length   20
NMissing 0

Fitted Trend Equation
Yt = 4140.63 + 27.1095*t

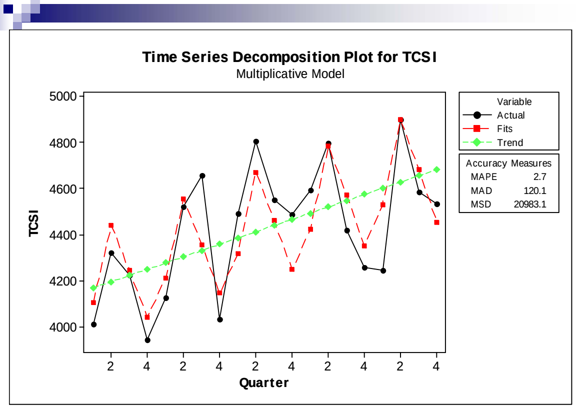
Seasonal Indices

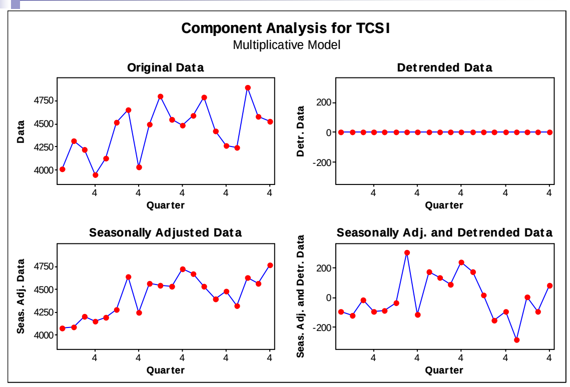
Period	Index
1	0.98469
2	1.05871
3	1.00536
4	0.95124

Accuracy Measures

MAPE	2.7
MAD	120.1
MSD	20983.1

Time	TCSI	Trend	Seasonal	Detrend	Deseason	Predict	Error
1	4009	4167.74	0.98469	0.96191	4071.33	4103.94	-94.938
2	4321	4194.85	1.05871	1.03007	4081.38	4441.13	-120.132
3	4224	4221.96	1.00536	1.00048	4201.48	4244.59	-20.588
4	3944	4249.07	0.95124	0.92820	4146.17	4041.88	-97.884
1	4123	4276.18	0.98469	0.96418	4187.10	4210.72	-87.716
2	4522	4303.29	1.05871	1.05082	4271.23	4555.94	-33.937
3	4657	4330.40	1.00536	1.07542	4632.17	4353.61	303.393
4	4030	4357.51	0.95124	0.92484	4236.58	4145.03	-115.034
1	4493	4384.62	0.98469	1.02472	4562.85	4317.49	175.506
2	4806	4411.73	1.05871	1.08937	4539.49	4670.74	135.259
3	4551	4438.84	1.00536	1.02527	4526.74	4462.63	88.373
4	4485	4465.95	0.95124	1.00427	4714.90	4248.18	236.815
1	4595	4493.06	0.98469	1.02269	4666.44	4424.27	170.728
2	4799	4520.17	1.05871	1.06169	4532.87	4785.55	13.454
3	4417	4547.28	1.00536	0.97135	4393.45	4571.65	-154.646
4	4258	4574.38	0.95124	0.93084	4476.27	4351.34	-93.335
1	4245	4601.49	0.98469	0.92253	4311.00	4531.05	-286.050
2	4900	4628.60	1.05871	1.05863	4628.27	4900.35	-0.350
3	4585	4655.71	1.00536	0.98481	4560.56	4680.66	-95.665
4	4533	4682.82	0.95124	0.96801	4765.36	4454.49	78.514





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* . *4*nd Ed. McGraw-Hill Companies

Terima kasih