filkom.ub.ac.id

DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics

Pemodelan Linier: Regresi dan Korelasi

Oleh: Imam Cholissodin | imamcs@ub.ac.id, Putra Pandu Adikara, Sufia Adha Putri
Asisten: Guedho, Sukma, Anshori, Aang dan Gusti
Fakultas Ilmu Komputer (Filkom) Universitas Brawijaya (UB)

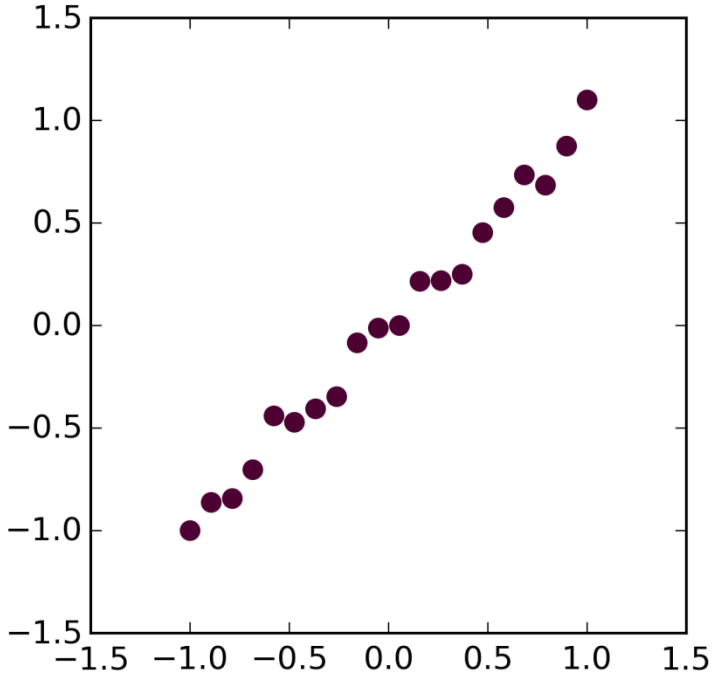
digitalent.kominfo.go.id

Pokok Pembahasan

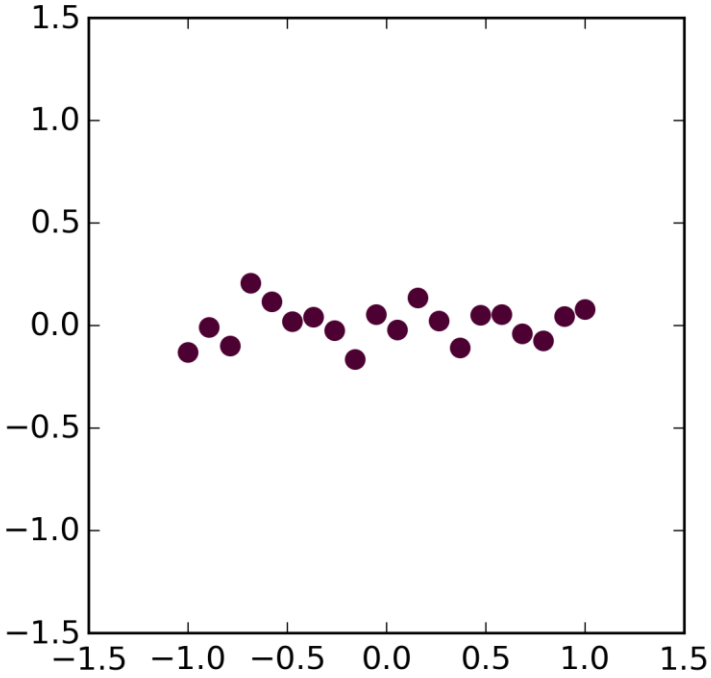
- Regresi – estimasi hubungan antar variabel
  - Regresi Linier
  - Pengujian asumsi
  - Regresi Non-linier
- Korelasi
  - Koefisien korelasi mengkuantifikasi kuatnya asosiasi antar variabel
  - Sensitivitas terhadap distribusi
- Tugas

Hubungan antar variabel

Terhubung

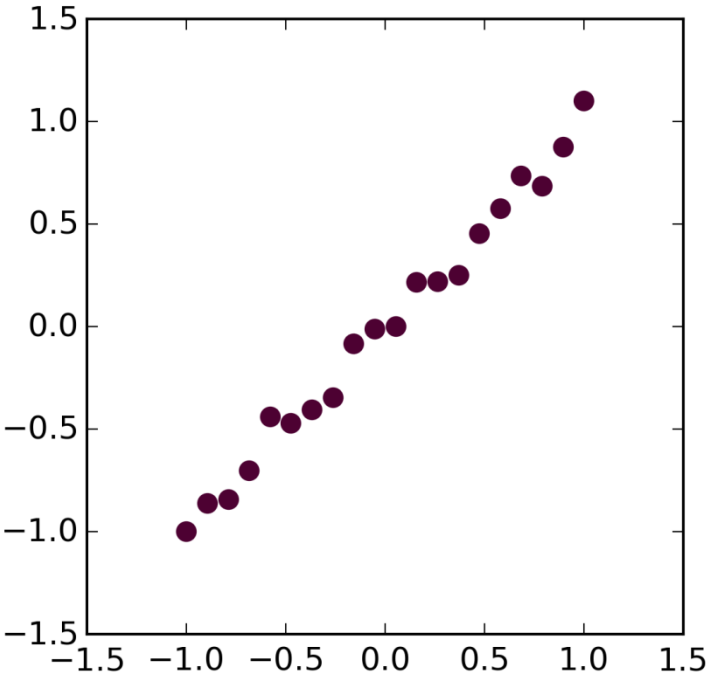


Tak Terhubung

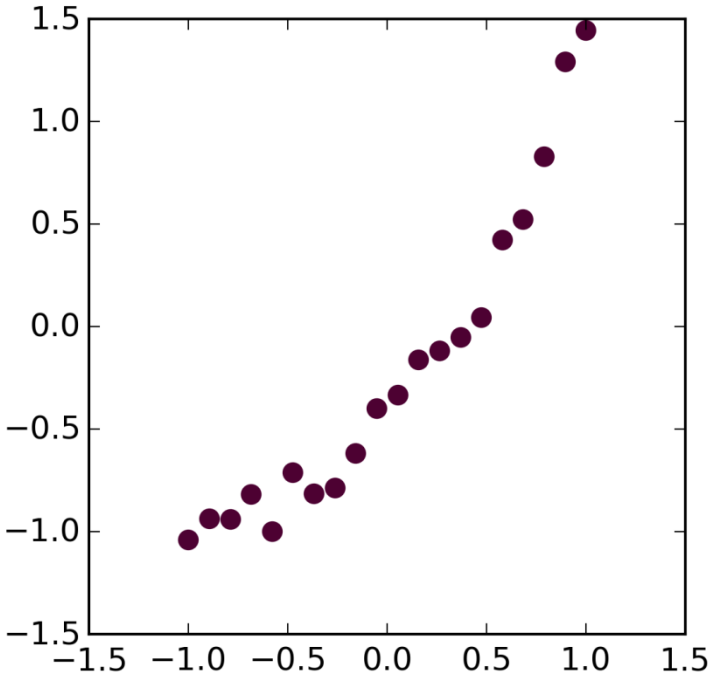


Hubungan antar variabel

Terhubung Linier

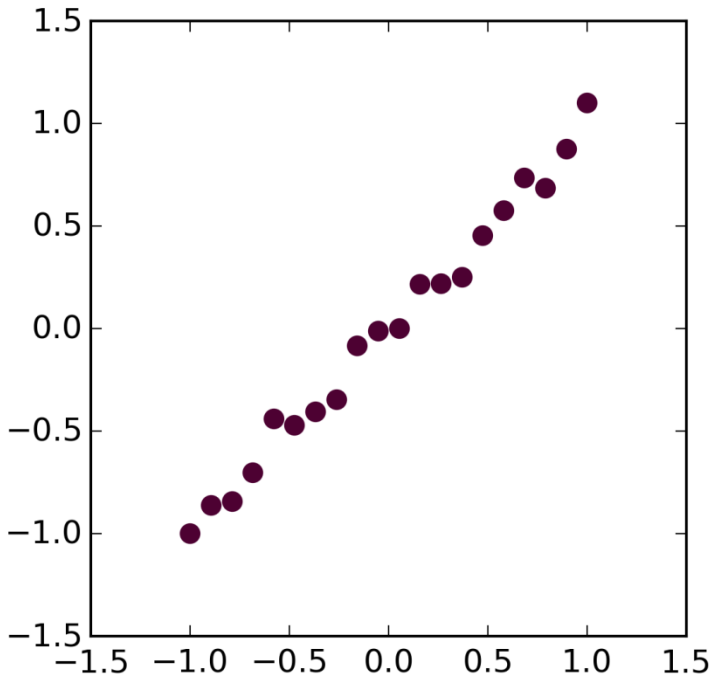


Terhubung Non-linier

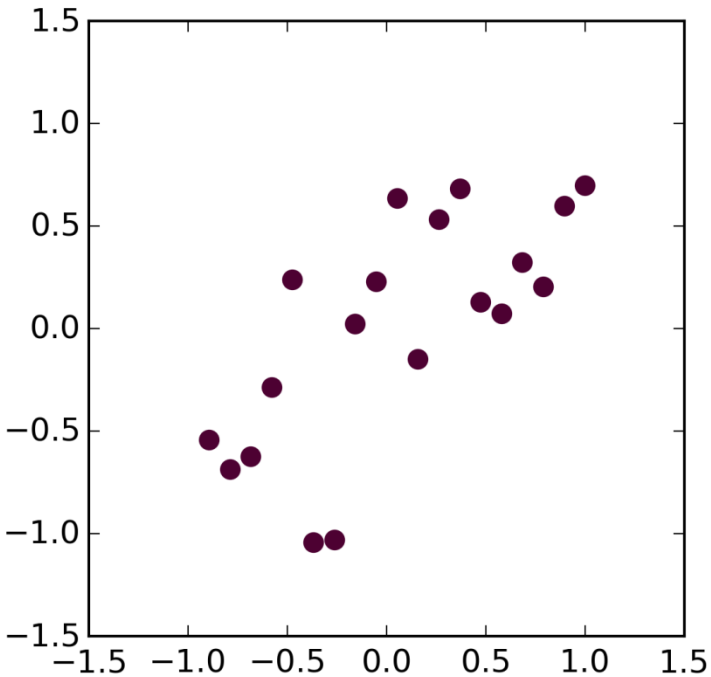


Hubungan antar variabel

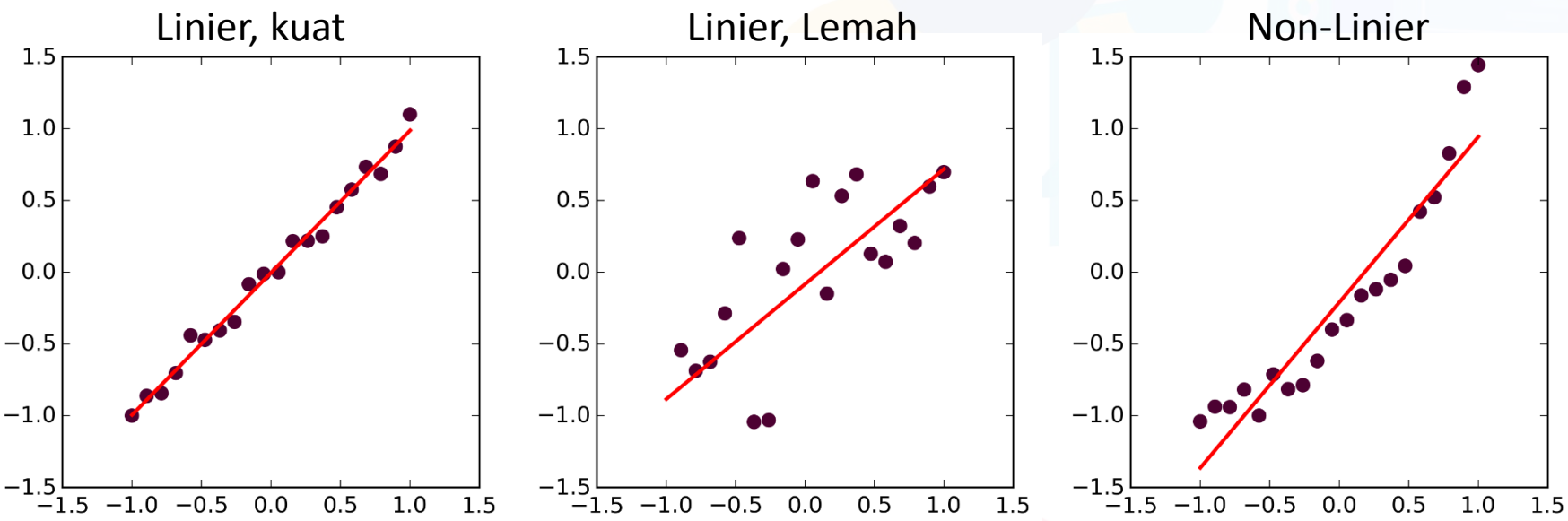
Linier, terhubung kuat



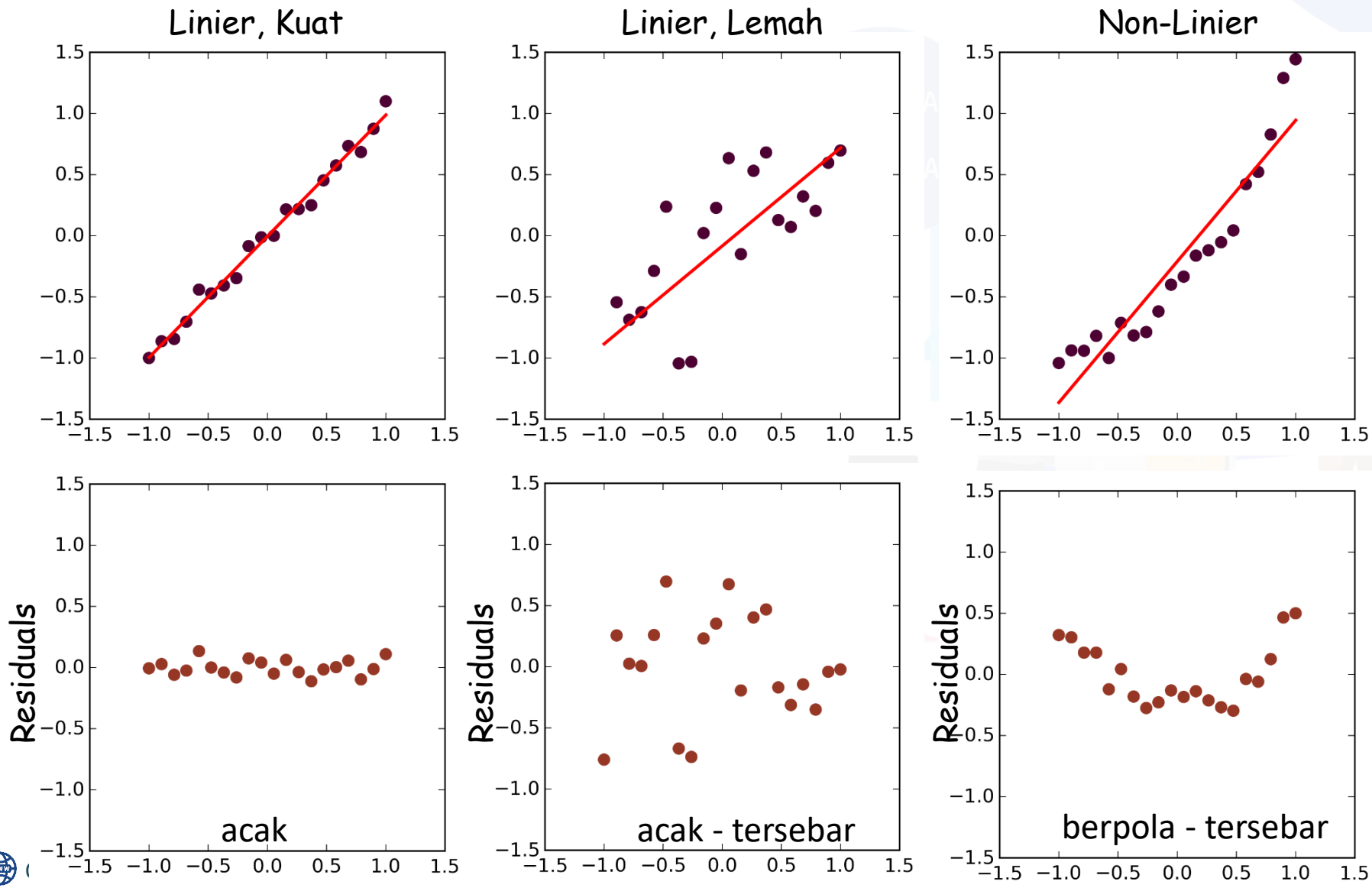
Linier, terhubung lemah



Regresi Linier



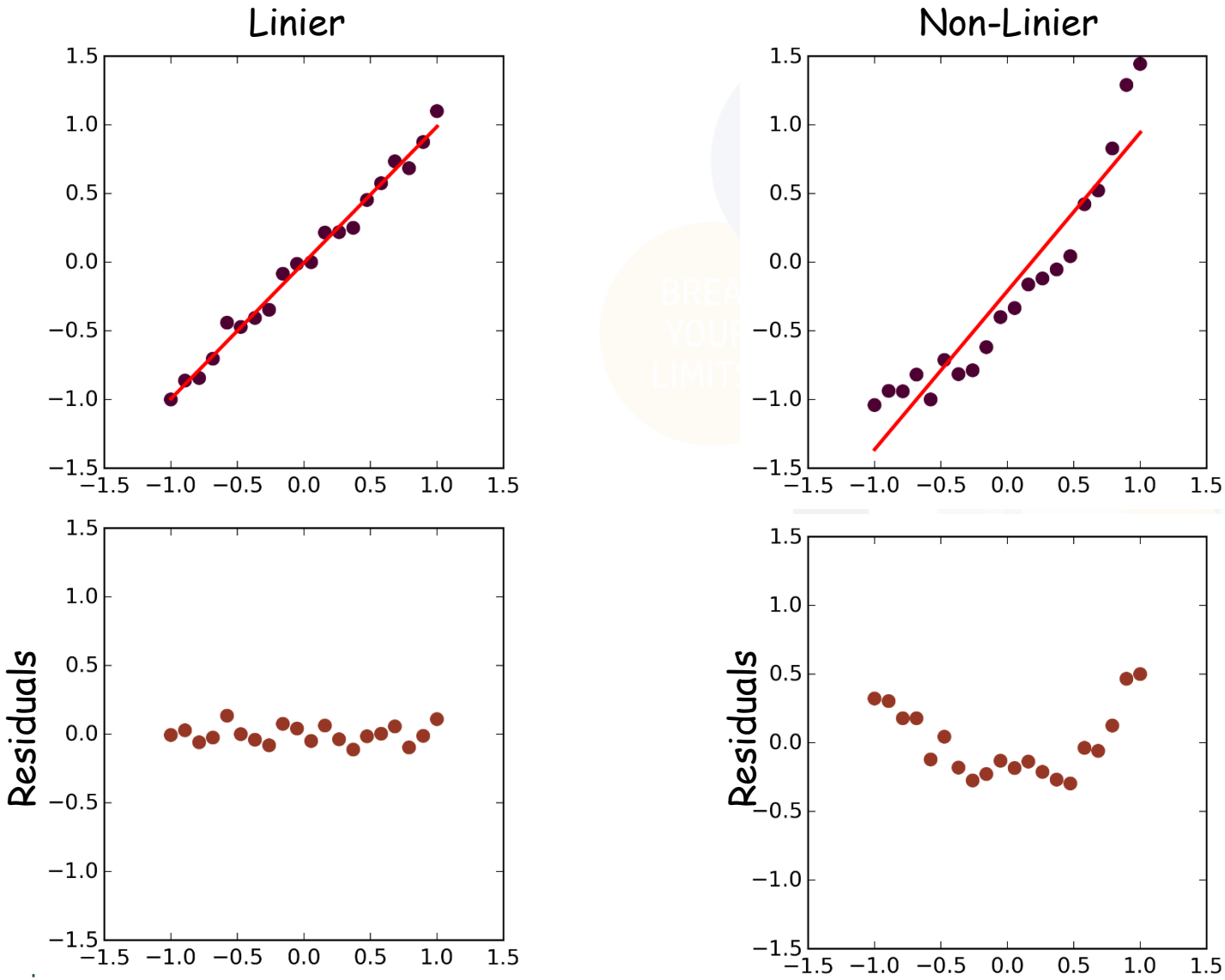
Regresi Linier - Residual



Asumsi-asumsi pada Regresi Linier

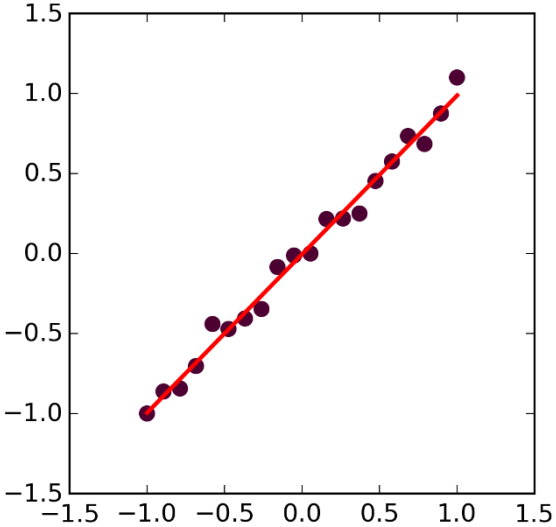
- Hubungan antar variabel adalah linier.
- Errors bersifat bebas, terdistribusi normal dengan nilai rata-rata nol dan varians yang konstan.

Asumsi-asumsi pada Regresi Linier

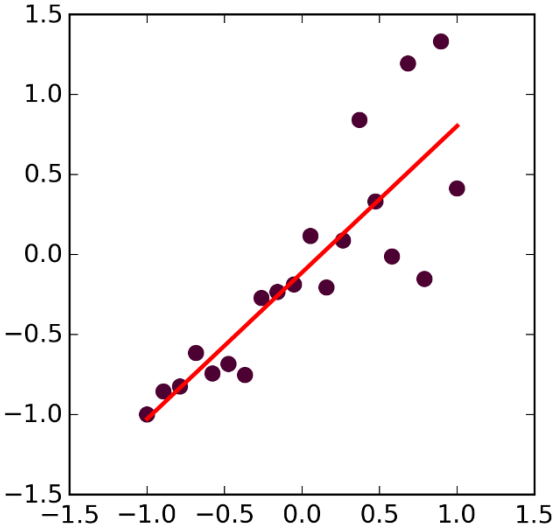


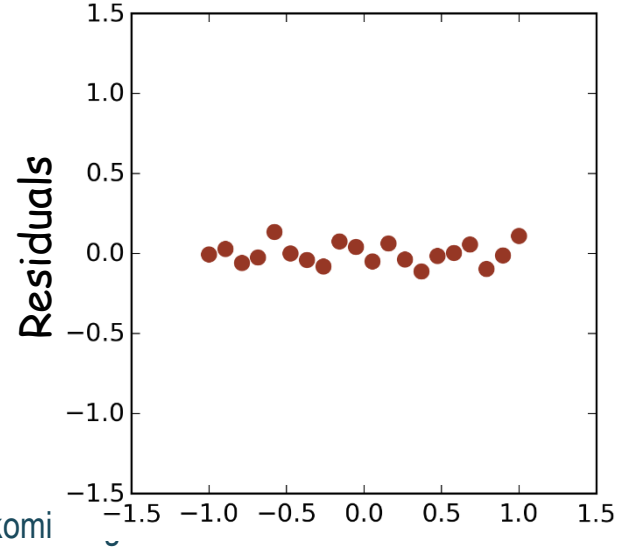
Asumsi-asumsi pada Regresi Linier

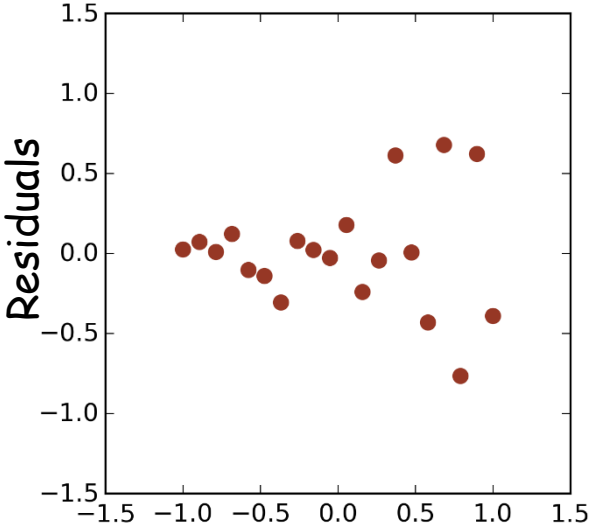
Varians konstan



Varians berubah-ubah







Model Regresi Linier

Intercept

Slope

Variabel bebas

Variabel terikat

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Komponen linier}} + \underbrace{\varepsilon_i}_{\text{Komponen random error}}$$

Random Error

Regresi Linier – Estimasi fungsi garis

Estimasi Intercept

Estimasi Slope

Estimasi Nilai Sampel

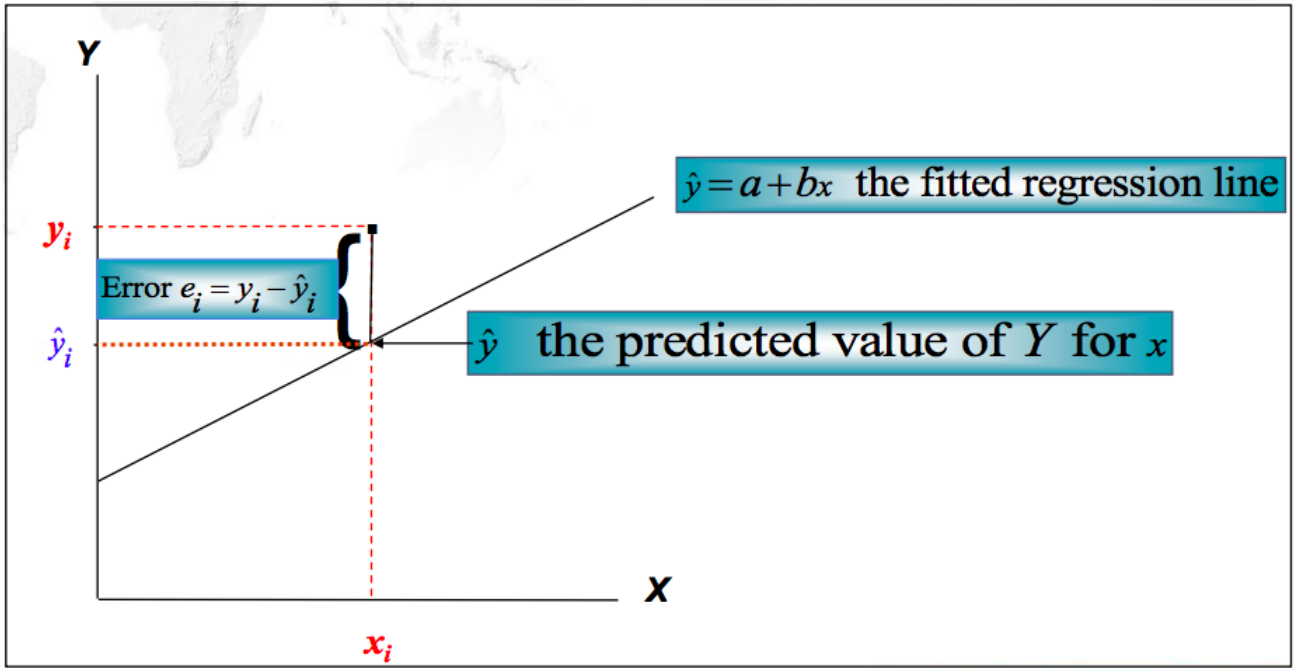
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Variabel Bebas

Metode *Least Squares*

Find slope and intercept given measurements X_i, Y_i, i=1..N that minimizes the **sum of the squares of the residuals**.

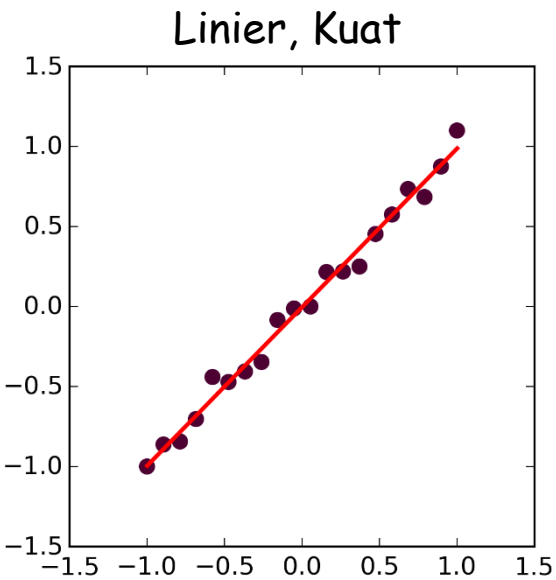
$$S = \sum \varepsilon_i^2$$

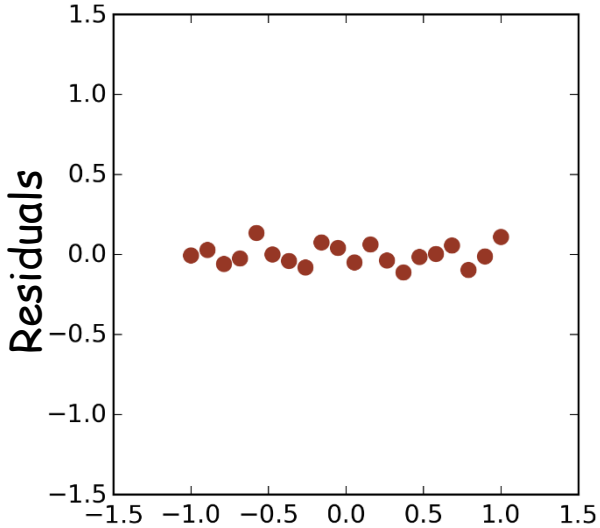


Linier Regresi in Python

```
import scipy.stats as stats

slope,intercept,r_value,p_value,std_err = stats.linregress(x,y)
```

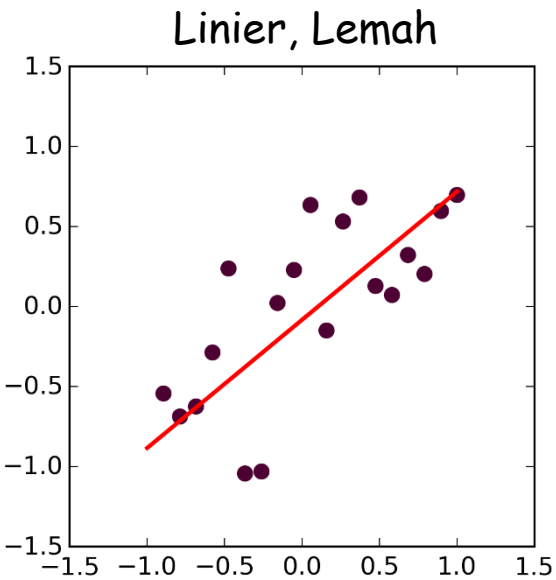
Linier Regresi Example



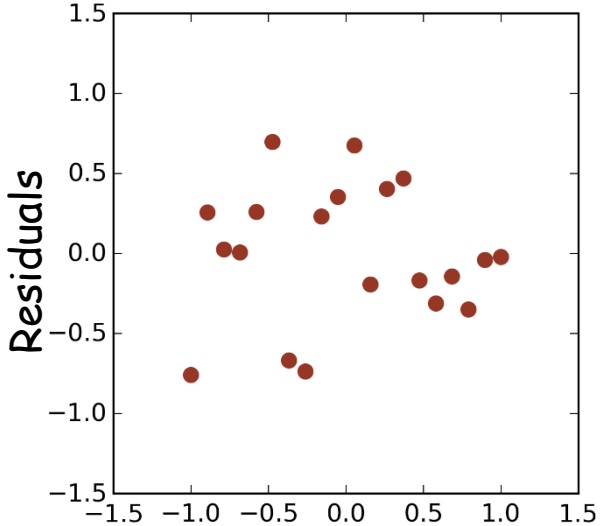



```
x=np.linspace(-1,1,points)
y=x+0.1*np.random.normal(size=points)
slope,intercept,r_value,p_value,std_err
= stats.linregress(x,y)
y_line=slope*x+intercept

fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y,color='#4D0132',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
ax1.plot(x,y_line,color='red',lw=2)
fig.savefig('Linier.png')

fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y-y_line,
color='#963725',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
fig.savefig('Linier-residuals.png')
```

Linier Regresi Example

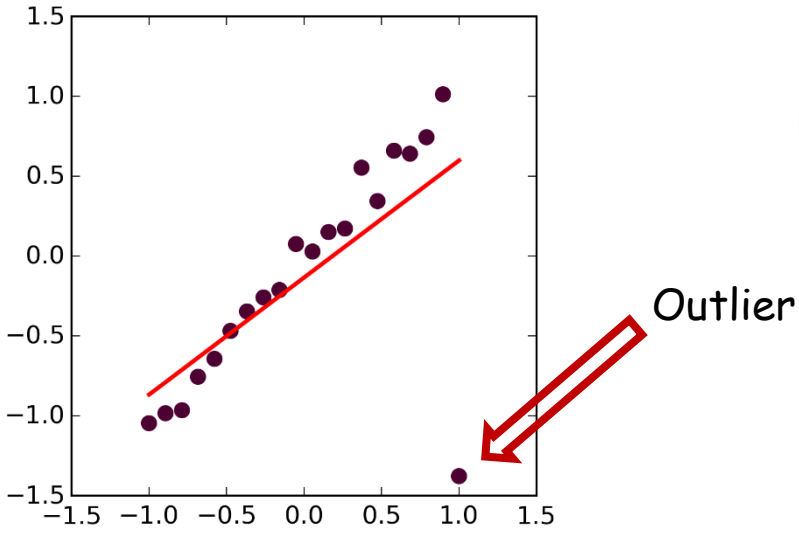


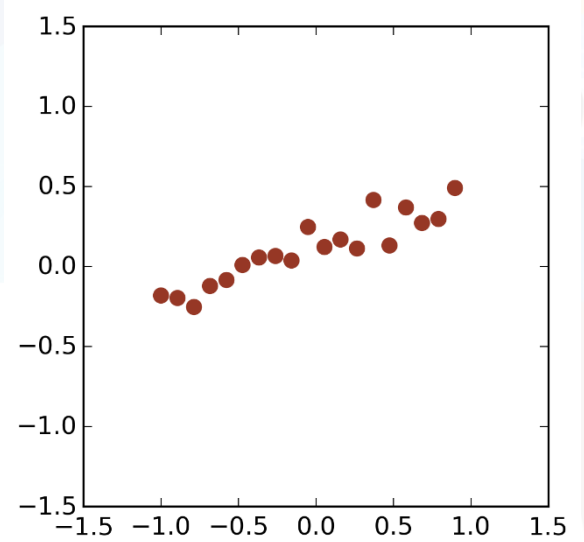

```
x=np.linspace(-1,1,points)
y=x+0.4*np.random.normal(size=points)
slope,intercept,r_value,p_value,std_err
= stats.linregress(x,y)
y_line=slope*x+intercept

fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y,color='#4D0132',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
ax1.plot(x,y_line,color='red',lw=2)
fig.savefig('Linier-Lemah.png')
```



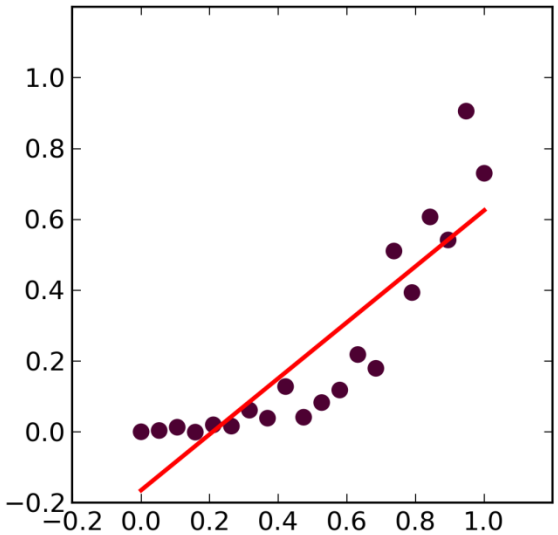

```
fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y-y_line,
color='#963725',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
fig.savefig('Linier-Lemah-residuals.png')
```

Linier Regresi Example

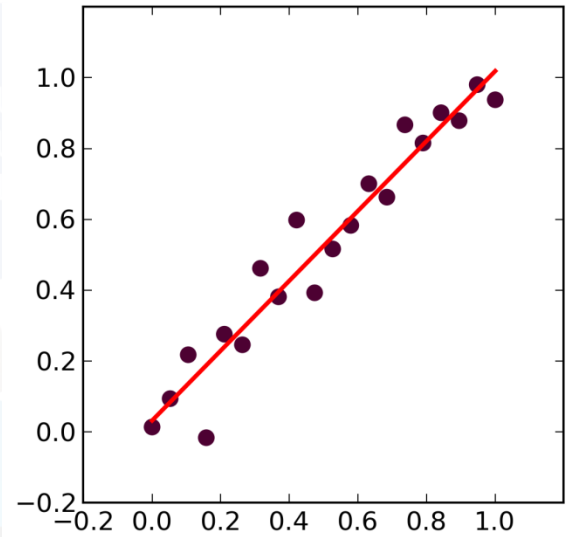




Regresi – Non-Linier data



Solution 1: Transformation



Solution 2: Non-Linier Regresi

$$\hat{Y}_i = f(X_i, \hat{\beta}_0, \hat{\beta}_1, \ldots)$$

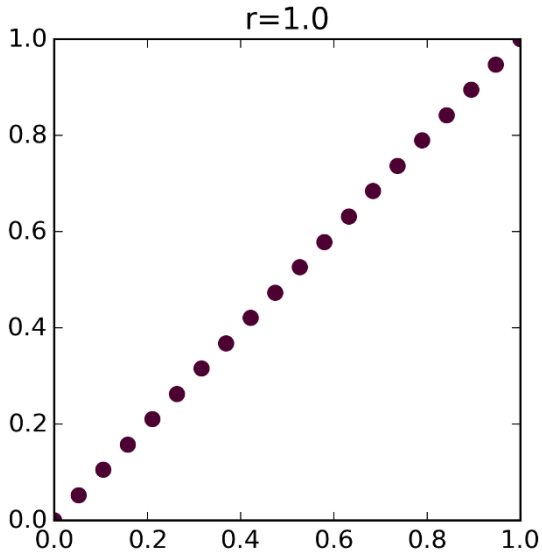
Koefisien Korelasi

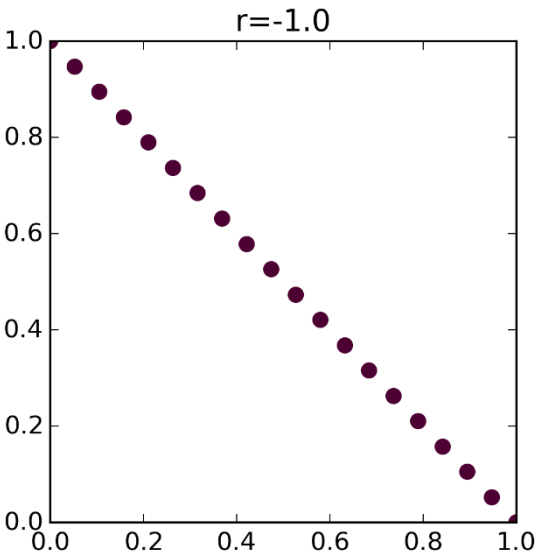
- A measure of the korelasi between the two variables
- Quantifies the association strength

Pearson korelasi coefficient:

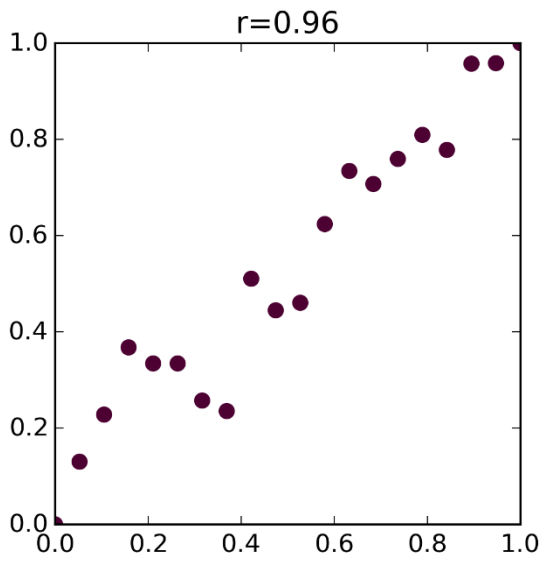
$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

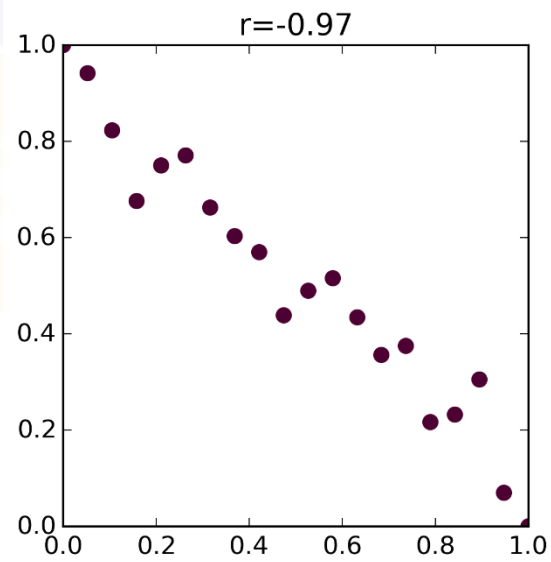
Koefisien Korelasi



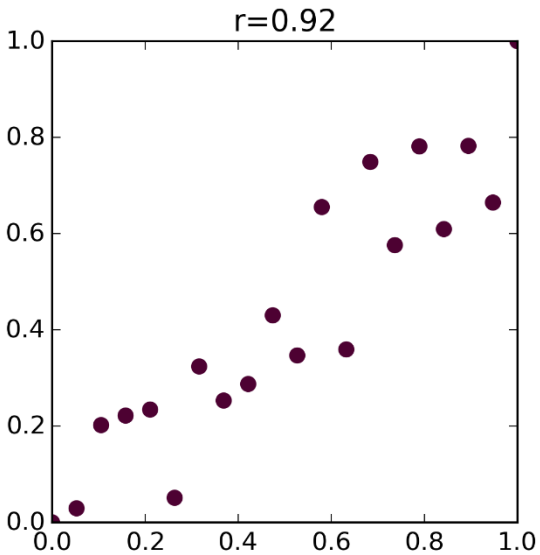


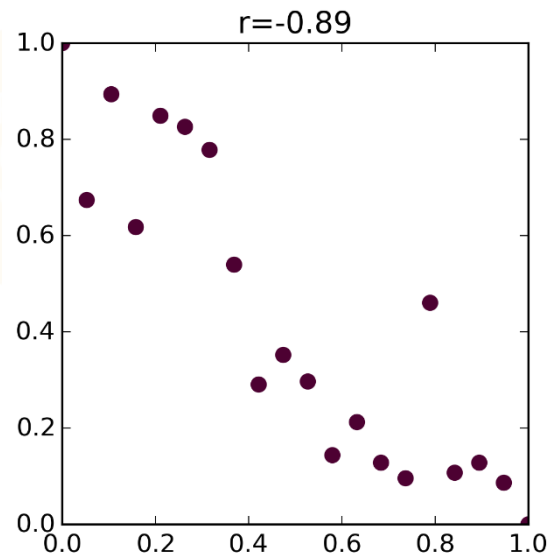
Koefisien Korelasi



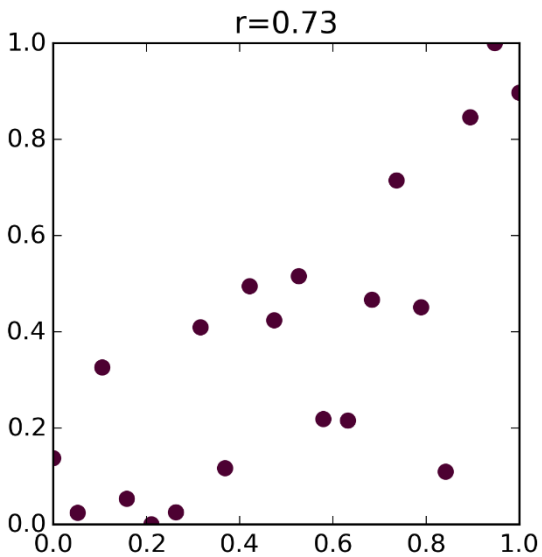


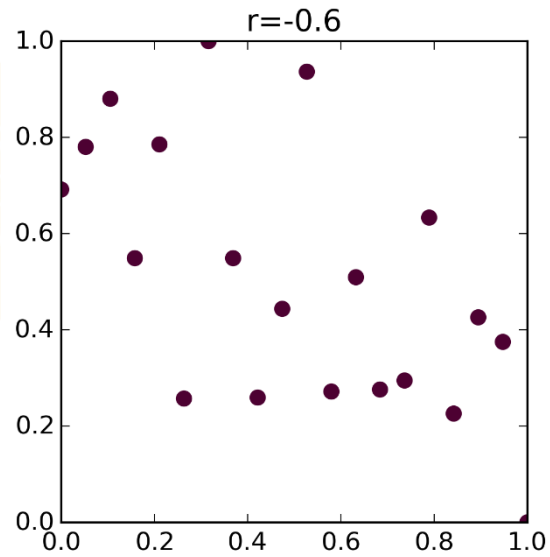
Koefisien Korelasi



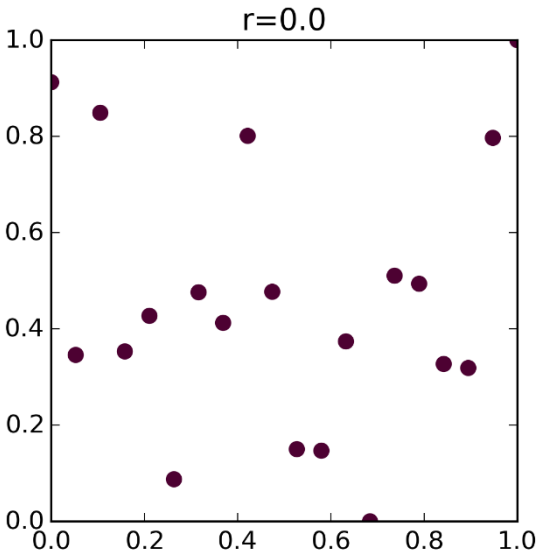


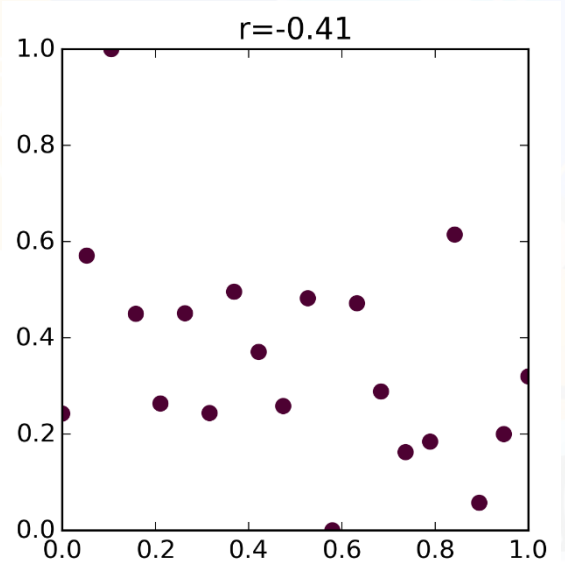
Koefisien Korelasi



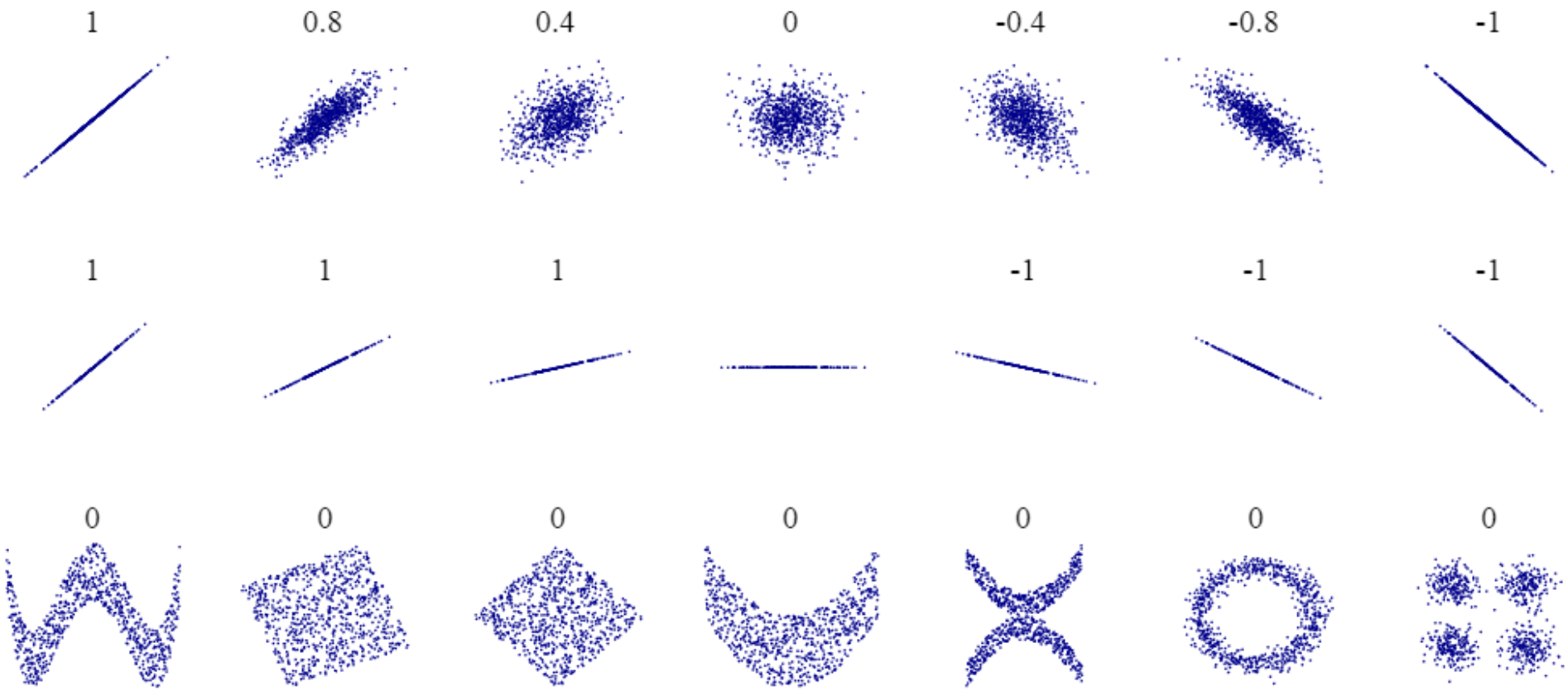


Koefisien Korelasi





Koefisien Korelasi



Source: Wikipedia

Koefisien Varians

Sample

$$X_1, X_2, \ldots, X_n$$

Mean

$$\mu = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

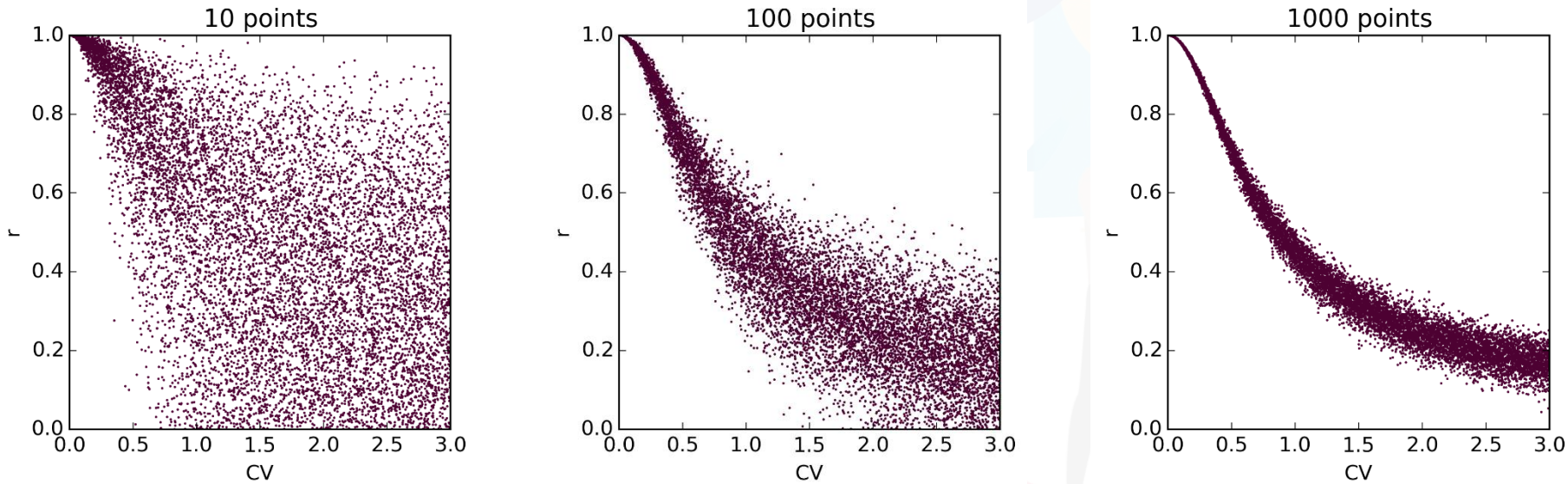
Variance

$$\sigma^2 = \frac{\sum_{i=1}^{i=n} (x_i - \mu)^2}{n}$$

Coefficient of Variation (CV) = $\dfrac{\sigma}{\mu}$

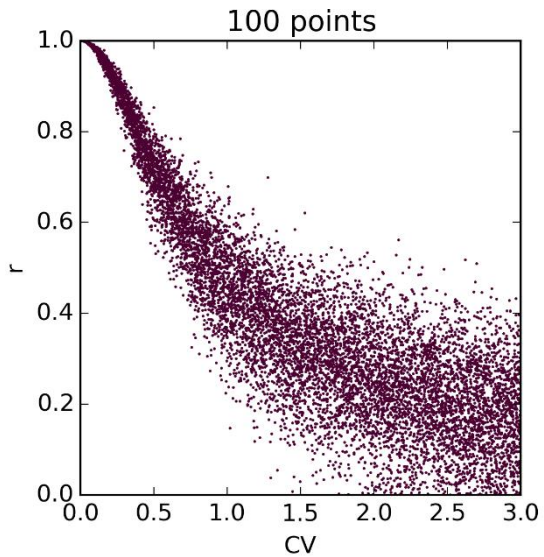
Koefisien Korelasi dan Koefisien Varians

Uniform distribusi

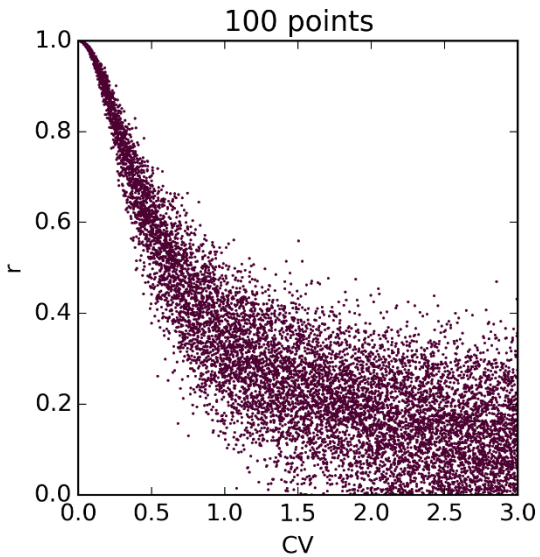


Koefisien Korelasi dan Koefisien Varians

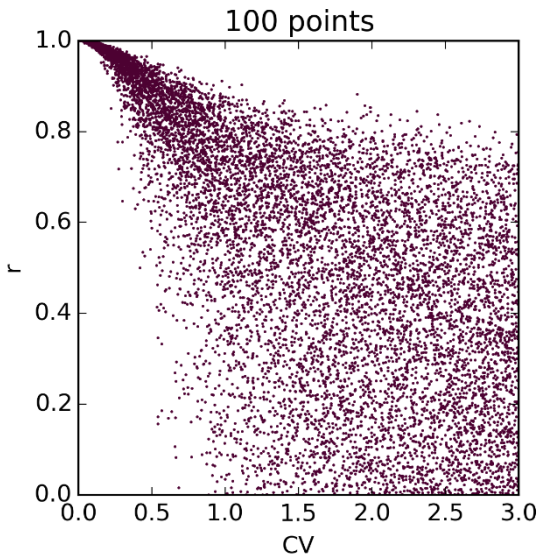
Uniform distribusi



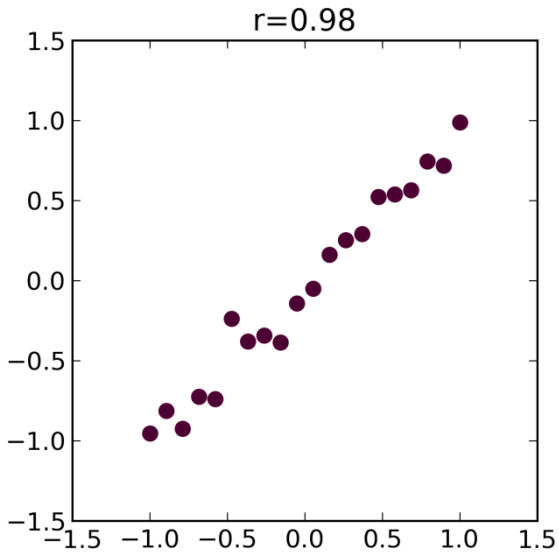
Normal distribusi

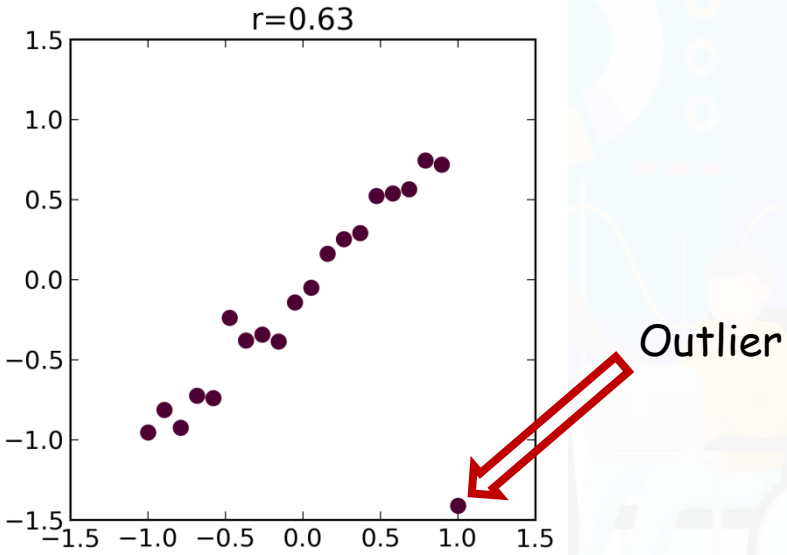


Lognormal distribusi

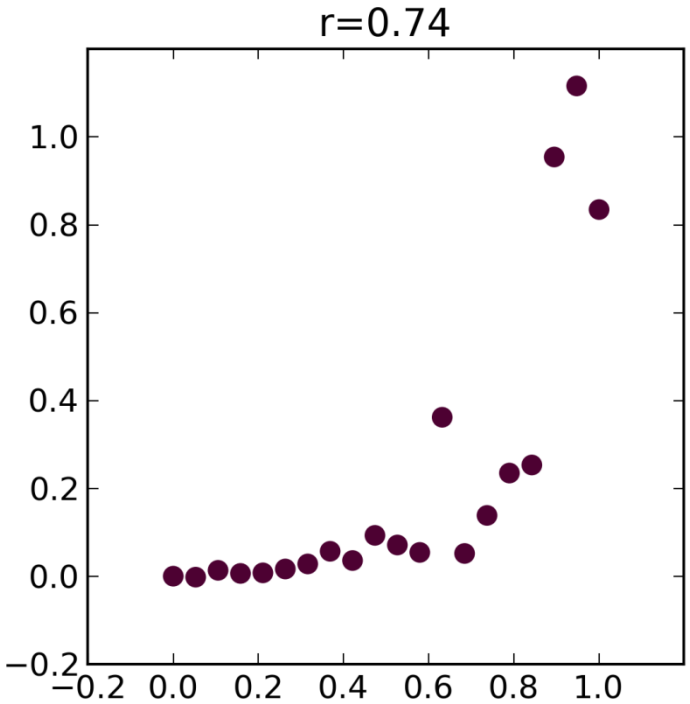


Koefisien Korelasi - Outliers





Koefisien Korelasi – Non Linier

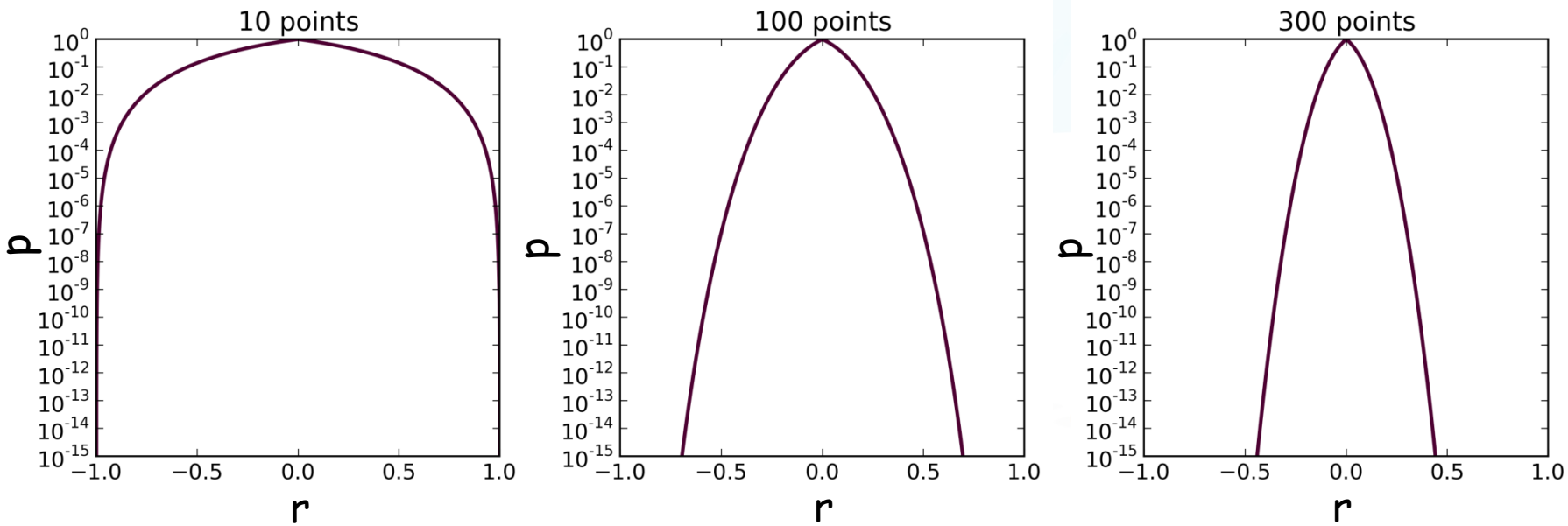


Solutions:

- Transformation
- Rank korelasi(Spearman, r=0.93)

Koefisien Korelasi dan p-Value

Hypothesis: Is there a correlation?



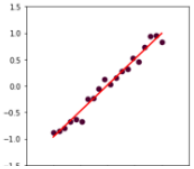
Latihan langsung di Kelas Ke-1 & Pembahasan Link kode “http://bit.ly/2ZBY7gp”

Silahkan dicoba dijalankan dengan Jupyter notebook yang Anda buat sebelumnya di Ubuntu 16.04 atau dengan SageMaker notebook (JupyterLab) yang baru Anda buat hari ini.

Lab-Sesi26-1

In [8]:
```
import scipy.stats as stats
import numpy as np
import matplotlib.pyplot as plt
```

In [11]:
```
# Linier Regresi Example Ke-1
points = 20
x=np.linspace(-1,1,points)
y=x+0.1*np.random.normal(size=points)
slope,intercept,r_value,p_value,std_err    = stats.linregress(x,y)
y_line=slope*x+intercept

fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y,color='#4D0132',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
ax1.plot(x,y_line,color='red',lw=2)
fig.savefig('Linier.png')

fig, (ax1) = plt.subplots(1,figsize=(4,4))
ax1.scatter(x,y-y_line, color='#963725',lw=0,s=60)
ax1.set_xlim([-1.5,1.5])
ax1.set_ylim([-1.5,1.5])
fig.savefig('Linier-residuals.png')

plt.show()
plt.close()
```



Latihan langsung di Kelas Ke-2 & Pembahasan

- Tidak ada

Tugas Individu

1. Buatlah rangkuman materi dengan cara berikut:
   - Menambahkan penjelasan/comment pada tiap program ipynb pada tugas latihan dari hasil “Latihan langsung di Kelas Ke-1” kedalam file *.doc/docx.

*semua bentuk tugas tersebut (merger dlm 1 file pdf) upload ke turnitin untuk cek plagiasi.

filkom.ub.ac.id

DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics

Terimakasih

Oleh: Imam Cholissodin | imamcs@ub.ac.id, Putra Pandu Adikara, Sufia Adha Putri
Asisten: Guedho, Sukma, Anshori, Aang dan Gusti
Fakultas Ilmu Komputer (Filkom) Universitas Brawijaya (UB)

digitalent.kominfo.go.id