

ANALISIS KORELASI & REGRESI

Kompilasi
Kelompok 6 dan 8

PERMASALAHAN

- Internal Revenue Service mencoba menduga pajak aktual yang tertunda setiap bulan dari divisi auditingnya. Diduga dua faktor yang mempengaruhi adalah jumlah jam kerja pegawai dan jumlah jam kerja mesin (komputer). Untuk menganalisis seberapa besar kedua faktor tersebut mempengaruhi besarnya pajak aktual yang tertunda (yang tidak dibayar) setiap bulan, dicatat pajak setiap variabel selama 10 bulan.

Responden	Y (RP 1000) (Pajak aktual yang tidak di bayar)	X1 (jam kerja pegawai)	X2 (jam kerja mesin/ komputer)
Januari	29	45	16
Februari	24	42	14
Maret	27	44	15
April	25	45	13
Mei	26	43	13
Juni	28	46	14
Juli	30	44	16
Agustus	28	45	16
September	28	44	15
Oktober	27	43	15

Uji Asumsi Pra Analisis Regresi

Uji Kenormalan

```
>shapiro.test(residuals(fm))
```

```
Shapiro-Wilk normality test
```

```
data: residuals(fm)
```

```
W = 0.94452, p-value = 0.6043
```

- ▶ Hipotesis
Ho : Residual berdistribusi normal
H1 : Residual tidakberdistribusi normal
- ▶ Tingkat signifikansi $\alpha = 0.05$
- ▶ Statistik Uji *Shapiro Wilk*
W = 0.94452
p - value = 0.6043
- ▶ Keputusan
Karena $\alpha = 0.05 < p - \text{value}$ maka Ho tidak ditolak (Ho diterima)
d.k.l asumsi kenormalan dianggap dapat dipenuhi

```
> cor(X1,X2)
```

```
[1] 0.1840943
```

```
> cor.test(X1,X2)
```

Pearson's product-moment correlation

data: X1 and X2

$t = 0.52975$, $df = 8$, $p\text{-value} = 0.6107$

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.5039445 0.7291987

sample estimates:

cor

0.1840943

► Menyusun Hipotesis

H_0 : Tidak ada korelasi ganda antara X1 dengan X2 ($\text{cor}(X1, X2) = 0$)

H_1 : Terdapat korelasi ganda X1 dengan X2 ($\text{cor}(X1, X2) \neq 0$)

► Tingkat signifikansi $\alpha = 0.05$

► Daerah kritis (daerah penolakan H_0)

$p\text{-value} = 0.6107$

► Keputusan

Karena $\alpha = 0.05 < p\text{-value} = 0.6107$ maka H_0 diterima

d.k.l tidak ada korelasi antara X1 dan X2

Korelasi Y dengan X1

```
> cor(Var.Y, Var.X1)
[1] 0.5015167
> cor.test(Var.Y, Var.X1)
```

Pearson's product-moment correlation

```
data: Var.Y and Var.X1
t = 1.6396, df = 8, p-value = 0.1397
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1872313  0.8596826
sample estimates:
      cor
0.5015167
```

► Menyusun Hipotesis

Ho : Tidak ada korelasi ganda antara X1 dengan Y ($\text{cor}(Y, X1) = 0$)

H1 : Terdapat korelasi ganda X1 dengan Y ($\text{cor}(Y, X1) \neq 0$)

► Tingkat signifikansi $\alpha = 0.05$

► Daerah kritis (daerah penolakan Ho)

p - value = 0.1397

► Keputusan

Karena $\alpha = 0.05 < p\text{-value} = 0.1397$
maka Ho diterima

d.k.l tidak ada korelasi antara Y dengan X1

Korelasi Y dengan X2

```
> cor(Var.Y,Var.X2)
[1] 0.7714616
> cor.test(Var.Y,Var.X2)
```

Pearson's product-moment correlation

```
data: Var.Y and Var.X2
t = 3.4294, df = 8, p-value = 0.008963
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2758006 0.9430284
sample estimates:
      cor
0.7714616
```

► Menyusun Hipotesis

Ho : Tidak ada korelasi ganda antara X2 dengan Y ($\text{cor}(Y, X2) = 0$)

H1 : Terdapat korelasi ganda X2 dengan Y ($\text{cor}(Y, X2) \neq 0$)

► Tingkat signifikansi $\alpha = 0.05$

► Daerah kritis (daerah penolakan Ho)
p - value = 0.008963

► Keputusan

Karena $\alpha = 0.05 > \text{p - value} = 0.008963$
maka Ho ditolak

d.k.l ada korelasi antara Y dengan X2

Korelasi Y dengan X1,X2

```
> cor(Var.Y,Var.X1+Var.X2)
[1] 0.8243243
> cor.test(Var.Y,Var.X1+Var.X2)
```

Pearson's product-moment correlation

```
data: Var.Y and Var.X1 + Var.X2
t = 4.1185, df = 8, p-value = 0.003351
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4047913 0.9571659
sample estimates:
cor
0.8243243
```

► Menyusun Hipotesis

Ho : Tidak ada korelasi ganda antara X1 dan X2 dengan Y ($\text{cor}(Y, X1+X2) = 0$)

H1 : Terdapat korelasi ganda antara X1 dan X2 dengan Y ($\text{cor}(Y, X1+X2) \neq 0$)

► Tingkat signifikansi $\alpha = 0.05$

► Daerah kritis (daerah penolakan Ho)

p - value = 0.003351

► Keputusan

Karena p - value = 0.003351 < $\alpha = 0.05$ maka Ho tidak diterima (Ho ditolak)

► Kesimpulan

Terdapat korelasi ganda antara X1 dan X2 dengan Y . Dengan kata lain, terdapat korelasi antara jumlah jam kerja pegawai dan jam kerja mesin terhadap besarnya pajak yang tertunda.

Uji Homogenitas

```
> var.test(Var.X1,Var.X2)
```

```
F test to compare two variances
```

```
data: Var.X1 and Var.X2
```

```
F = 1.0661, num df = 9, denom df = 9, p-value = 0.9256
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.2648081 4.2921756
```

```
sample estimates:
```

```
ratio of variances
```

```
1.066116
```

- ▶ Menyusun hipotesis

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 \neq \sigma_2$

- ▶ Tingkat signifikansi

- ▶ Statistik uji $\alpha = 0.05$

- ▶ Dari perhitungan yang diperoleh dari *software R*

p - value = 0.9256

- ▶ Keputusan

Karena $\alpha = 0.05 < p\text{-value} = 0.9256$
maka H_0 tidak ditolak (H_0 diterima)

- ▶ Kesimpulan

Variansi-variansi dari kedua populasi tersebut sama (homogen).

ANALISIS REGRESI

```
> fm<-lm(Y~X1+X2)
```

```
> summary(fm)
```

Call:

```
lm(formula = Y ~ X1 + X2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.24668	-0.74702	-0.02321	0.51956	1.42706

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-13.8196	13.3233	-1.037	0.33411
X1	0.5637	0.3033	1.859	0.10543
X2	1.0995	0.3131	3.511	0.00984 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.071 on 7 degrees of freedom

Multiple R-squared: 0.7289, Adjusted R-squared: 0

1. Bentuk Model Persamaan Regresi

$$Y = -13.8196 + 0.5637 X_1 + 1.0995 X_2$$

3. Uji signifikansi masing-masing koefisien

- Perhatikan bahwa koefisien X1 tidak signifikan
- Berbeda dengan X2!!!
- Kaitkan dengan analisis korelasi antara Y dengan X1

4. Sebanyak 72.89% X1 dan X2 bisa menerangkan Y, sisanya 27.11% diterangkan variabel lain yang tidak masuk model

> anova(fm)

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	7.4450	7.4450	6.495	0.038186 *
X2	1	14.1312	14.1312	12.328	0.009844 **
Residuals	7	8.0239	1.1463		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2. Uji signifikansi Model (model langkah 1)

Pada tingkat kepercayaan 95%, model Signifikan, namun bila diambil tingkat Kepercayaan 99% model menjadi tidak signifikan

PEMERIKSAAN SISA

- Menghitung residual terstandar dengan fungsi rsstandard

```
> sres<-rstandard(fm)
```

```
> sres[1:10]
```

1	2	3	4	5	6
-0.14973264	-1.57144439	-0.46861538	-1.06428595	1.53821347	0.64108799
7	8	9	10		
1.53842167	-1.24583375	0.52112131	0.09506472		

- Mencari DFBETAS

```
> dfb<-dfbetas(fm)
```

```
> head(dfb)
```

	(Intercept)	X1	X2
1	0.047703673	-0.03013471	-0.05430691
2	-1.464895245	1.35977696	0.23240092
3	-0.009608251	0.02077722	-0.04342257
4	0.222044696	-0.50691942	0.79613421
5	0.851722296	-0.49206007	-0.98353502
6	-0.391626465	0.48491900	-0.25582131

- Residual terbesar ada pada observasi ke-2

- Mencari DFBETAS

Pada X1 data ke 2 memiliki DFBETAS sebesar 1.35977696 > 1, maka observasi tersebut dianggap mencurigakan.

► Mencari dengan DFFITS

```
> dff<-dffits(fm)
```

```
> dff[1:10]
```

1	2	3	4	5	6
-0.08527224	-1.63884494	-0.15451423	-0.99111248	1.39151975	0.57858447
7	8	9	10		
1.00897044	-0.80295127	0.17248903	0.04607053		

► Jarak Cook's

```
> cooksD<-cooks.distance(fm)
```

```
> cooksD[1:10]
```

1	2	3	4	5	6
0.0028186923	0.6760134747	0.0089933140	0.3201926362	0.4984858567	0.1225408686
7	8	9	10		
0.2620420210	0.1951348610	0.0111215249	0.0008243485		

- Mencari dengan DFFITS
- Observasi dicurigai jika nilai DFFITS >1. Pada analisis tersebut ditemukan nilai DFFITS > 1 yaitu pada data ke 5 (1.39151975) dan ke 7 (1.00897044).
- Jarak Cook's
- Observasi dicurigai jika nilai Jarak Cook's >1. Pada analisis tersebut tidak ditemukan nilai Jarak Cook's > 1