

PORTOFOLIO

Assa Trissia Rizal

Kata Pengantar Penulis



assatrissiarzl@gmail.com



Assa Trissia Rizal



Assa Trissia Rizal



@aassatz



0895377154926

Nama saya Assa Trissia Rizal. Kebanyakan orang memanggilku Assa. Saya lulusan dari Universitas Tanjungpura yang lulus pada tahun 2021, Jurusan Matematika Program Studi Statistika. Melalui portofolio ini, saya ingin menunjukkan apa yang telah saya lakukan sejauh ini dalam proyek berbasis data *science*. Mudah-mudahan, pada akhir portofolio ini, Anda akan mendapatkan wawasan berharga dan di dapat mendiskusikan kemungkinan peluang-peluang yang ada. Selamat membaca!

Menu



Python & Microsoft Excel

Analisis Perilaku dan Profil Pelanggan di Sebuah Bank



SPSS & R

Analisis Efek Perbedaan Pakan Ayam Terhadap
Pertumbuhan Berat Badan Anak Ayam



SPSS & R

Analisis Serangga Mati Terhadap Dampak Uji Coba 6
Macam Insektisida

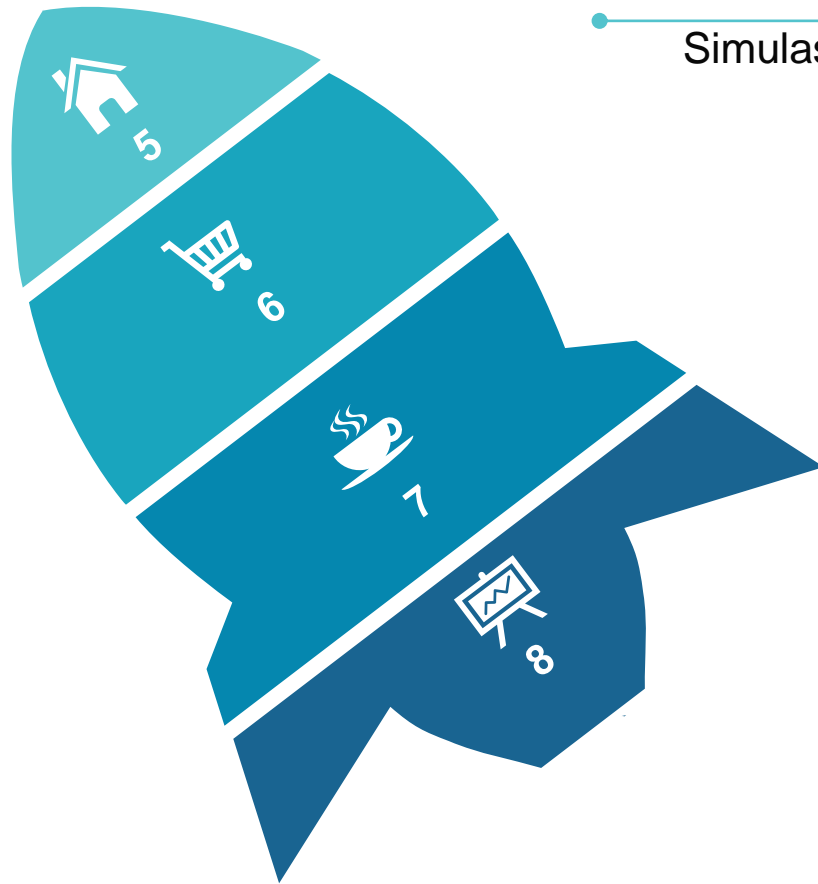


R Studio

Analisis Pengaruh Dosis Vitamin C Terhadap Pertumbuhan
Gigi Pada Hewan Marmut



Menu



Microsoft Excel

Simulasi Pendanaan Program Pensiun Manfaat Pasti dengan Metode
Spreading Gains and Losses



SPSS

Pengaruh Konsumsi Daging Sapi dan Daging Ayam
Terhadap Tingkat Permintaan Daging Perkapita di Amerika
Serikat




KNIME

Regresi Linear Pada Data Deret Waktu Kasus
Kejahatan yang terjadi di Amerika Serikat periode tahun
1997–2016



Sertifikat-Sertifikat





Analisis Perilaku dan Profil Pelanggan di Sebuah Bank

Aplikasi: Python & Microsoft Excel

Dataset ini memberikan informasi mengenai perilaku dan profil pelanggan di sebuah bank yang menyediakan suatu program pinjaman untuk pembelian properti. Data berisi semua data pribadi nasabah dan sekaligus rekam jejak kreditnya dalam satu periode.

loan_id	birth_date	phone_number	gender	married	dependents	education	self_employed	applicant_income	coapplicant_income	loan_amount
LP001002	5/15/1977	628114203187.00	Male	No	0	Graduate	No	5849	0	100000
LP001003	8/24/1979	628776449212.00	Male	Yes	1	Graduate	No	4583	1508	128000
LP001005	3/28/1991	628119240537.00	Male	Yes	0	Graduate	Yes	3000	0	66000
LP001006	6/2/1980	62856343702.00	Male	Yes	0	Not Graduate	No	2583	2358	120000
LP001008	5/19/1989	628113008687.00	Male	No	0	Graduate	No	6000	0	141000
LP001011	1/22/1986	62811028825.00	Male	Yes	2	Graduate	Yes	5417	4196	267000
LP001013	6/16/1979	628111840127.00	Male	Yes	0	Not Graduate	No	2333	1516	95000
LP001014	5/21/1999	628567566695.00	Male	Yes	3+	Graduate	No	3036	2504	158000
LP001018	8/25/1987	628528959314.00	Male	Yes	2	Graduate	No	4006	1526	168000
LP001020	12/13/1983	628118659695.00	Male	Yes	1	Graduate	No	12841	10968	349000
LP001024	3/8/1984	628116652331.00	Male	Yes	2	Graduate	No	3200	700	70000
LP001027	5/6/1983	628110331279.00	Male	Yes	2	Graduate		2500	1840	109000
LP001028	11/16/1987	628112778301.00	Male	Yes	2	Graduate	No	3073	8106	200000
LP001029	6/13/1976	628132105871.00	Male	No	0	Graduate	No	1853	2840	114000
LP001030	9/4/1986	628115076013.00	Male	Yes	2	Graduate	No	1299	1086	17000
LP001032	9/7/1989	628774718199.00	Male	No	0	Graduate	No	4950	0	125000
LP001034	11/17/1987		Male	No	1	Not Graduate	No	3596	0	100000
LP001036	6/25/1976	628561626867.00	Female	No	0	Graduate	No	3510	0	76000
LP001038		628112236123.00	Male	Yes	0	Not Graduate	No	4887	0	133000
LP001041	4/19/1992	628112135606.00	Male	Yes	0	Graduate		2600	3500	115000
LP001043	12/11/1993	62857672414.00	Male	Yes	0	Not Graduate	No	7660	0	104000

Dari dataset yang telah ada, bahwa banyak data yang tidak ada terisi. Sehingga harus diatasi yaitu dengan menggunakan *missing value* pada masing-masing kolom. Dari analisis yang dilakukan bahwa *missing value* terbanyak terjadi pada variabel *has_credit_card*. Selanjutnya dilakukan analisis dengan mengecek *duplicated* data. Setalal *duplicated* data dapat diatasi, maka langkah selanjutnya adalah *handling outliers*. *Outliers* adalah data point (baris) yang nilainya ekstrim/jauh berbeda dari data-data lain pada umumnya. *Outliers* dapat muncul karena kesalahan pada pengambilan data dan keberadaan individu-individu yang 'spesial'. *Outliers* harus diatasi karena dapat menyebabkan model *machine learning* berperforma buruk. Terakhir yaitu saya melakukan *Feature Encoding*. *Feature Encoding* adalah proses mengubah *feature categorical* menjadi *feature numeric*. Label Encoding adalah perubahan feature categorical menjadi numeric dengan memberikan angka yang berbeda bagi masing-masing nilai unik.

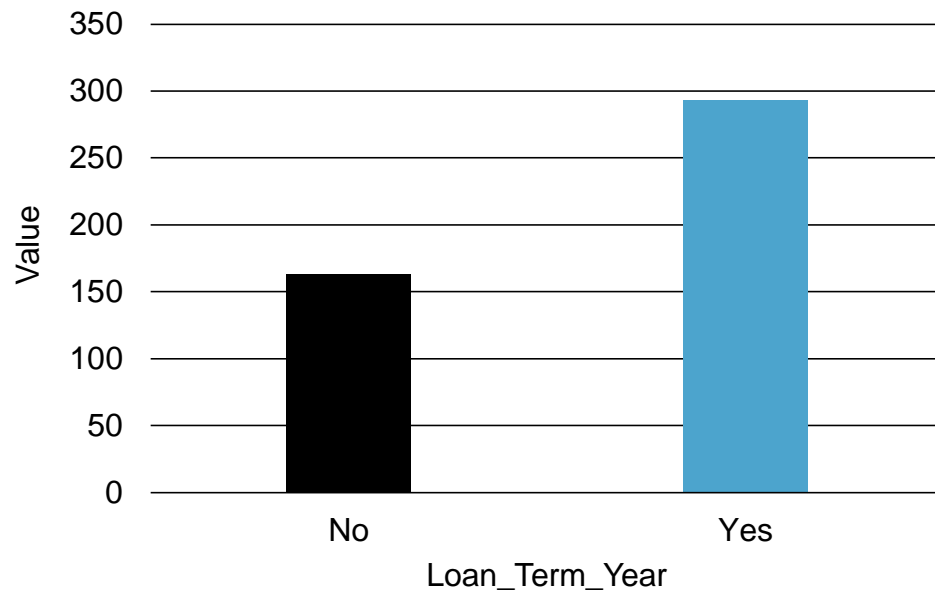


Gender	Property Type		
	Property Type	Property Type	Property Type
	House	Studio	Apartment
Female	43	25	21
Male	134	114	119



Dari visualisasi yang ditampilkan memuat informasi data gender dan data *Property Type*. Data gender terdiri dari *Female* dan *Male*. Sedangkan data *property type* terdiri dari *Property Type House*, *Property Type Studio*, dan *Property Type Apartment*. Dari hasil tabel dapat dianalisis bahwa *Male* lebih banyak mengajukan sebuah pinjaman di bank dibandingkan dengan *Female*. Pembelian *property* terbanyak baik *Female* maupun *Male* yaitu *Property Type Apartment*. Pembelian *Property Type Studio* paling sedikit di beli *Female*. *Property Type House* adalah *property* paling sedikit di beli *Male*.

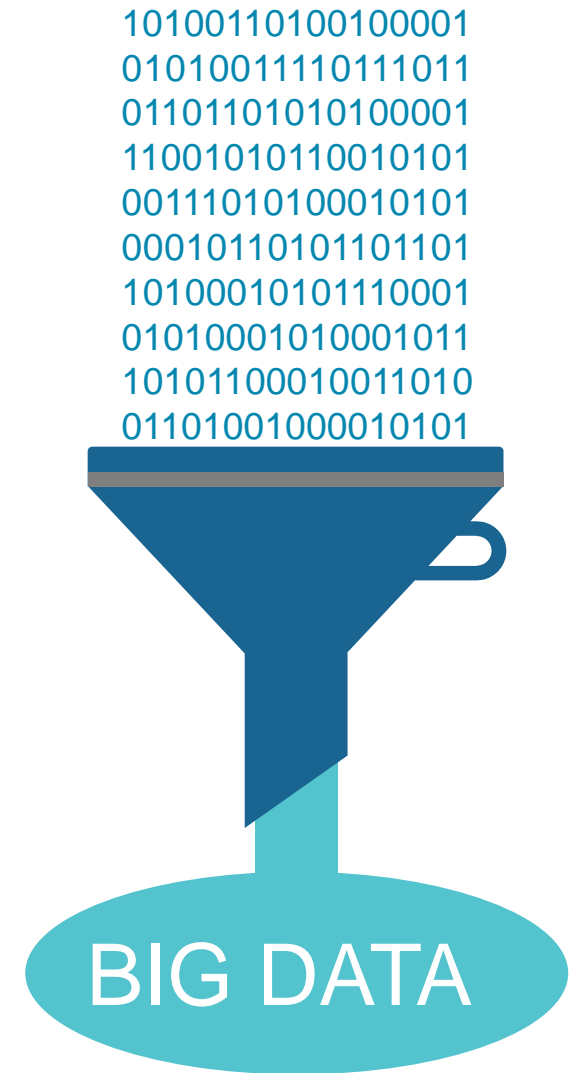
	Loan_term_year	
	Married	Value
0	No	163
1	Yes	293



Dari visualisasi yang ditampilkan, bahwa status perkawinan yang sudah menikah lebih banyak mengambil pinjaman dalam pembelian properti yaitu sebanyak 293 nasabah. Sedangkan yang belum menikah lebih sedikit mengambil pinjaman dalam pembelian properti yaitu sebanyak 163 nasabah. Maka dapat disimpulkan bahwa status pernikahan mempengaruhi pengambilan pinjaman untuk pembelian sebuah properti.

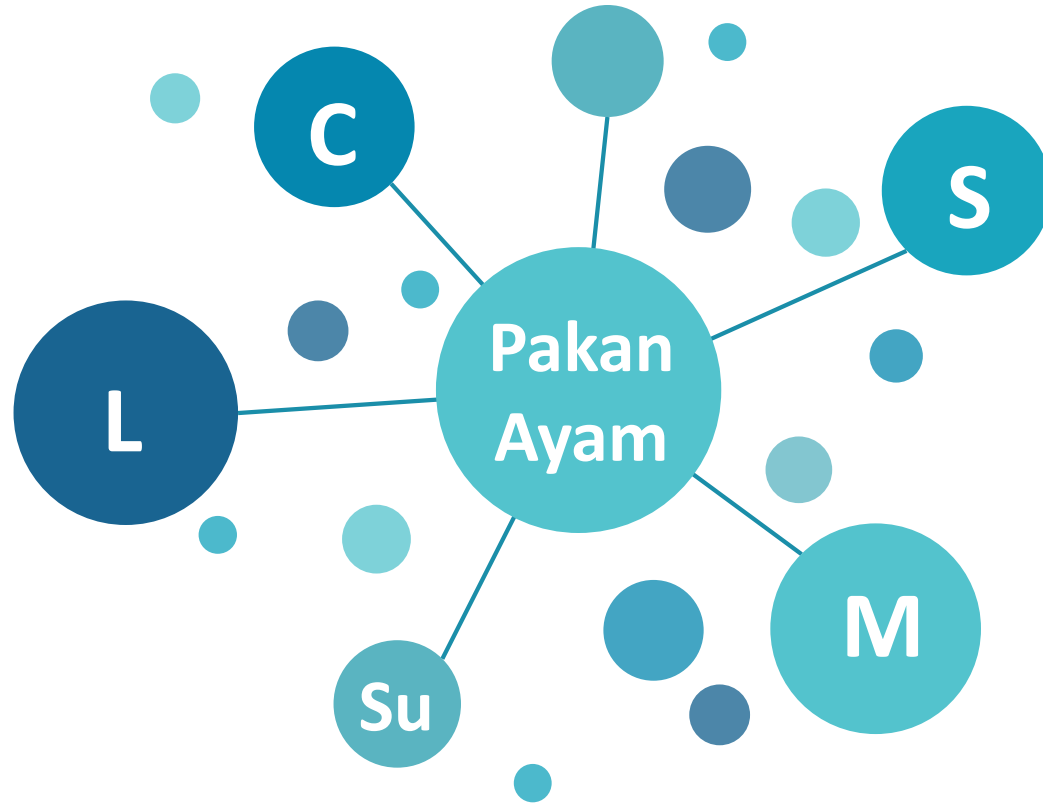
<i>Married</i>	<i>Gender</i>	<i>Loan Status</i>	<i>Loan Term Year</i>	<i>Property Type Apartement</i>	<i>Property Type House</i>	<i>Property Type Studio</i>
<i>No</i>	<i>Female</i>	APPROVE	43	43	43	43
		REJECT	22	22	22	22
	<i>Male</i>	APPROVE	69	69	69	69
		REJECT	29	29	29	29
<i>Yes</i>	<i>Female</i>	APPROVE	17	17	17	17
		REJECT	7	7	7	7
	<i>Male</i>	APPROVE	194	194	194	194
		REJECT	75	75	75	75

Berdasarkan *section 2* dan *3* yang telah dianalisis menggunakan *Microsoft Excel*, dapat disimpulkan bahwa pada status belum menikah maupun sudah menikah baik *Female* maupun *Male* selalu memiliki *loan status approve* lebih tinggi dibandingkan *reject*. Secara keseluruhan untuk yang berstatus belum menikah dengan gender *Female*, lebih banyak mengajukan pinjaman dibandingkan sudah menikah. Berbeda halnya dengan Gender *Male*, yang berstatus sudah menikah lebih banyak mengajukan pinjaman ke bank. Pembelian *property* terbanyak yaitu pada tipe apartement kemudian *house* dan yang paling sedikit adalah tipe studio. Sehingga dapat disimpulkan bahwa status pernikahan dan gender sangat mempengaruhi pengambilan pinjam seorang nasabah.



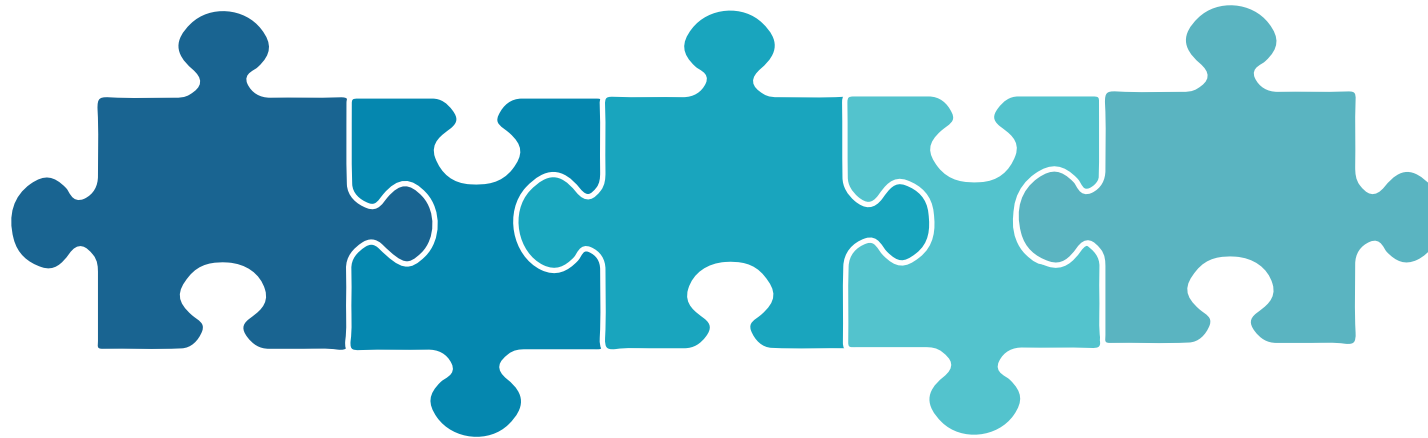
Analisis Efek Perbedaan Pakan Ayam Terhadap Pertumbuhan Berat Badan Anak Ayam

Aplikasi : SPSS dan R



Peneliti ingin meneliti, apakah terdapat efek perbedaan pakan ayam terhadap pertumbuhan berat badan anak ayam. Hipotesis awal dinyatakan bahwa tidak terdapat perbedaan pertumbuhan berat badan anak ayam pada semua kelompok pemberian pakan ayam. Dengan tingkat signifikansinya sebesar 5%.

Sebelum dilakukan uji anova dilakukan pengujian asumsi terhadap kenormalan untuk tiap grup dan homogenitas varian dari datanya. Uji normalitas menggunakan uji Saphiro-Walk test dan uji homogenitas bisa menggunakan uji Bartlett.



Output Pada SPSS

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Casein	.170	12	.200 [*]	.917	12	.259

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Linseed	.102	12	.200 [*]	.969	12	.903

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Soybean	.118	14	.200 [*]	.946	14	.506

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Meatmeal	.139	10	.200 [*]	.979	10	.961

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Sunflower	.236	12	.064	.928	12	.360

a. Lilliefors Significance Correction

Syntak dan Output pada R

#One-way Analysis of Variance

```
data("chickwts")
```

```
chick <- chickwts
```

```
by (data=chick$weight, INDICES=chick$feed, FUN= shapiro.test)
```

Chick\$feed: casein

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.91663, p-value = 0.2592

chick\$feed: horsebean

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.93758, p-value = 0.5264

chick\$feed: linseed

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.96931, p-value = 0.9035

chick\$feed: meatmeal

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.97914, p-value = 0.9612

chick\$feed: soybean

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.9464, p-value = 0.5064

chick\$feed sunflower

Shapiro-Wilk normality test

```
data: dd[x, ]
```

W = 0.92809, p-value = 0.3603

tes Homogenitas variance

```
Bartlett.test(weight~feed, data=chick)
```

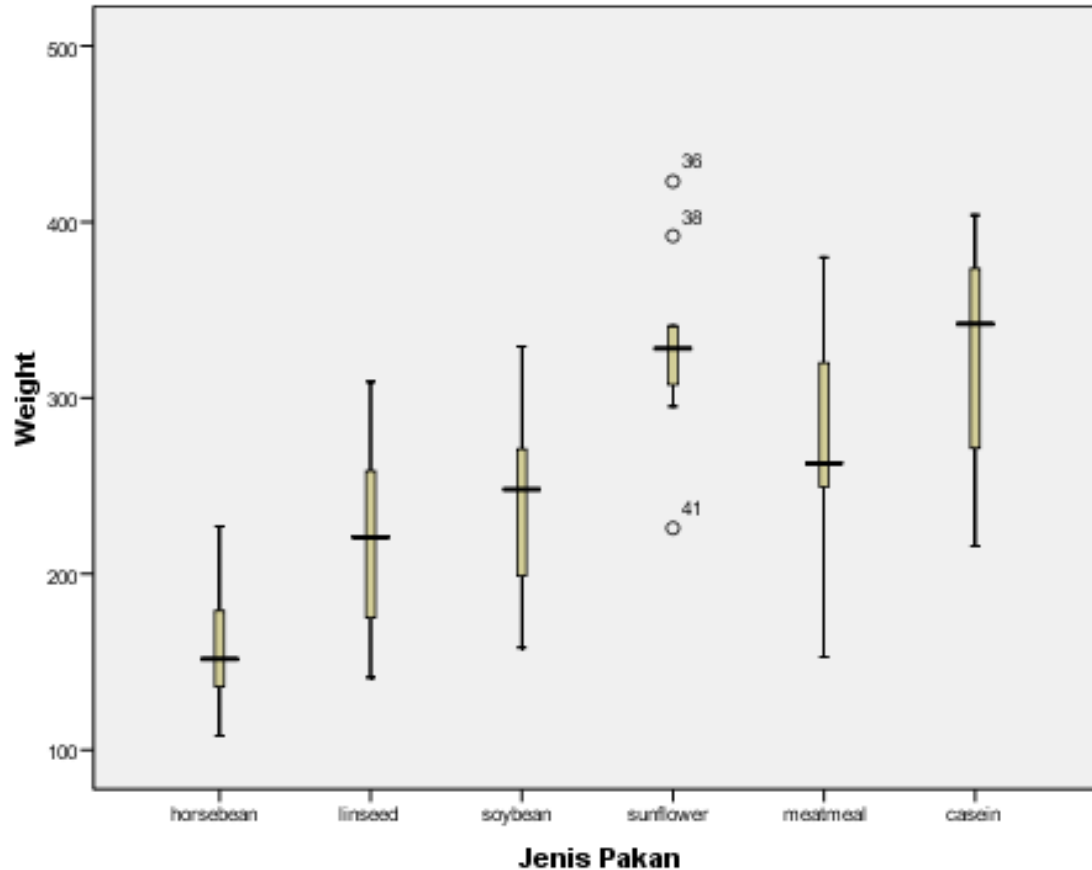
Bartlett test of homogeneity of variances

```
data: weight by feed
```

Bartlett's K-squared = 3.2597, df = 5, p-value = 0.66

Berdasarkan output bisa disimpulkan bahwa variabel yang akan dianalisis (*weight*) terdistribusi normal pada setiap grup pakan (feed) $p\text{-values} > \alpha=0.05$. Kemudian dari hasil uji Bartlett, terlihat bahwa ragam/varian *weight* dari grup pakan adalah homogen (bartlett.test $p\text{-value}=0.66$), sehingga bisa dilanjutkan dengan uji anova satu arah.





Dari hasil penelitian maka terdapat perbedaan dari efek pakan ayam terhadap berat pertumbuhan anak ayam dapat dilihat menggunakan box plot. Terlihat pada Gambar tinggi masing-masing box plot tiap jenis pakan (*feed*) berbeda posisi mediannya. Terlihat ayam dengan pakan horsebean memiliki berat badan yang paling kecil dibandingkan pakan lainnya. Sedangkan pakan Casein memberikan berat badan ayam yang lebih besar dibandingkan yang lainnya.

Box Plot antara variabel *weight* terhadap variabel *feed*

Multiple Comparisons						
Weight Bonferroni						
(I) Jenis Pakan	(J) Jenis Pakan	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
horsebean	linseed	-58.550	23.667	.240	-130.72	13.62
	soybean	-86.492*	23.250	.006	-157.39	-15.59
	sunflower	-168.717*	23.667	.000	-240.89	-96.55
	meatmeal	-116.709*	24.152	.000	-190.36	-43.06
	casein	-163.383*	23.667	.000	-235.55	-91.21
linseed	horsebean	58.550	23.667	.240	-13.62	130.72
	soybean	-27.942	22.128	1.000	-95.42	39.53
	sunflower	-110.167*	22.566	.000	-178.98	-41.35
	meatmeal	-58.159	23.073	.213	-128.52	12.20
	casein	-104.833*	22.566	.000	-173.65	-36.02
soybean	horsebean	86.492*	23.250	.006	15.59	157.39
	linseed	27.942	22.128	1.000	-39.53	95.42
	sunflower	-82.224*	22.128	.006	-149.70	-14.75
	meatmeal	-30.217	22.645	1.000	-99.27	38.84
	casein	-76.891*	22.128	.014	-144.37	-9.42
sunflower	horsebean	168.717*	23.667	.000	96.55	240.89
	linseed	110.167*	22.566	.000	41.35	178.98
	soybean	82.224*	22.128	.006	14.75	149.70
	meatmeal	52.008	23.073	.414	-18.35	122.37
	casein	5.333	22.566	1.000	-63.48	74.15
meatmeal	horsebean	116.709*	24.152	.000	43.06	190.36
	linseed	58.159	23.073	.213	-12.20	128.52
	soybean	30.217	22.645	1.000	-38.84	99.27
	sunflower	-52.008	23.073	.414	-122.37	18.35
	casein	-46.674	23.073	.709	-117.03	23.68
casein	horsebean	163.383*	23.667	.000	91.21	235.55
	linseed	104.833*	22.566	.000	36.02	173.65
	soybean	76.891*	22.128	.014	9.42	144.37
	sunflower	-5.333	22.566	1.000	-74.15	63.48
	meatmeal	46.674	23.073	.709	-23.68	117.03
*. The mean difference is significant at the 0.05 level.						

Output yang akan diperhatikan dalam uji ini adalah nilai *p-value* nya. Hipotesis awalnya adalah dalam tiap pasangan jenis kelompok feed ini tidak terdapat perbedaan. Dengan tingkat signifikansi dianggap 5%, maka jika terdapat nilai *p-value* yang lebih kecil dari 5% dinyatakan terdapat perbedaan yang signifikan antara dua kelompok tersebut.

Dari output tersebut setidaknya terdapat dua jenis kesimpulan dari masing-masing perbandingan kelompok, yaitu: berbeda secara signifikan dan tidak berbeda secara signifikan. Kelompok yang tidak berbeda secara signifikan adalah perbandingan efek faktor antar pakan: casein dengan meatmeal, casein dengan sunflower, sunflower dengan meatmeal, meatmeal dengan linseed, meatmeal dengan soybean, dan horsebean dengan linseed serta linseed dengan soybean.

Sedangkan pasangan kelompok yang berbeda secara signifikan adalah efek faktor antara pakan: casein dengan horsebean, casein dengan linseed, casein dengan soybean, horsebean dengan meatmeal, horsebean dengan soybean, horsebean dengan sunflower, linseed dengan sunflower dan soybean dengan sunflower.

Multiple Comparisons							
Dependent Variable:Weight							
	(I) Jenis Pakan	(J) Jenis Pakan	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	horsebean	linseed	-58.550	23.667	.147	-128.08	10.98
		soybean	-86.492*	23.250	.005	-154.80	-18.19
		sunflower	-168.717*	23.667	.000	-238.25	-99.19
		meatmeal	-116.709*	24.152	.000	-187.66	-45.76
		casein	-163.383*	23.667	.000	-232.91	-93.85
	linseed	horsebean	58.550	23.667	.147	-10.98	128.08
		soybean	-27.942	22.128	.804	-92.95	37.07
		sunflower	-110.167*	22.566	.000	-176.46	-43.87
		meatmeal	-58.159	23.073	.133	-125.94	9.63
		casein	-104.833*	22.566	.000	-171.13	-38.54
	soybean	horsebean	86.492*	23.250	.005	18.19	154.80
		linseed	27.942	22.128	.804	-37.07	92.95
		sunflower	-82.224*	22.128	.005	-147.23	-17.22
		meatmeal	-30.217	22.645	.765	-96.74	36.31
		casein	-76.891*	22.128	.011	-141.90	-11.88
	sunflower	horsebean	168.717*	23.667	.000	99.19	238.25
		linseed	110.167*	22.566	.000	43.87	176.46
		soybean	82.224*	22.128	.005	17.22	147.23
		meatmeal	52.008	23.073	.228	-15.78	119.79
		casein	5.333	22.566	1.000	-60.06	71.63

meatmeal	horsebean	116.709*	24.152	.000	45.76	187.66
	linseed	58.159	23.073	.133	-9.63	125.94
	soybean	30.217	22.645	.765	-36.31	96.74
	sunflower	-52.008	23.073	.228	-119.79	15.78
	casein	-46.674	23.073	.341	-114.46	21.11
casein	horsebean	163.383*	23.667	.000	93.85	232.91
	linseed	104.833*	22.566	.000	38.54	171.13
	soybean	76.891*	22.128	.011	11.88	141.90
	sunflower	-5.333	22.566	1.000	-71.63	60.96
	meatmeal	46.674	23.073	.341	-21.11	114.46
*. The mean difference is significant at the 0.05 level.						

Uji ini menggunakan tingkat signifikansi sebesar 5%. Output function juga menampilkan hasil pembuatan anova pada langkah sebelumnya. Terlihat bahwa formula yang dilakukan pada proses pembuatan anova test adalah antara variabel “weight” sebagai variabel “feed” sebagai variabel independen/faktor atau perlakuannya. Hasil perbandingan kelompok pakan (*feed*) dapat dijelaskan bahwa yang tidak terdapat perbedaan secara statistik ($p \text{ ajd} > 5\%$) adalah antara kelompok pakan *meatmeal-casien*, *sunflower-casien*, *linseed-horsebean*, *meatmeal-linseed*, *soybean-linseed*, *soybean-meatmeal* dan *sunflower-meatmeal*. Selain hasil perbandingan antara kedua kelompok tersebut adalah berbeda secara signifikan dalam mempengaruhi berat badan anak ayam.





Analisis Serangga Mati Terhadap Dampak Uji Coba 6 Macam Insektisida

Aplikasi: SPSS dan R

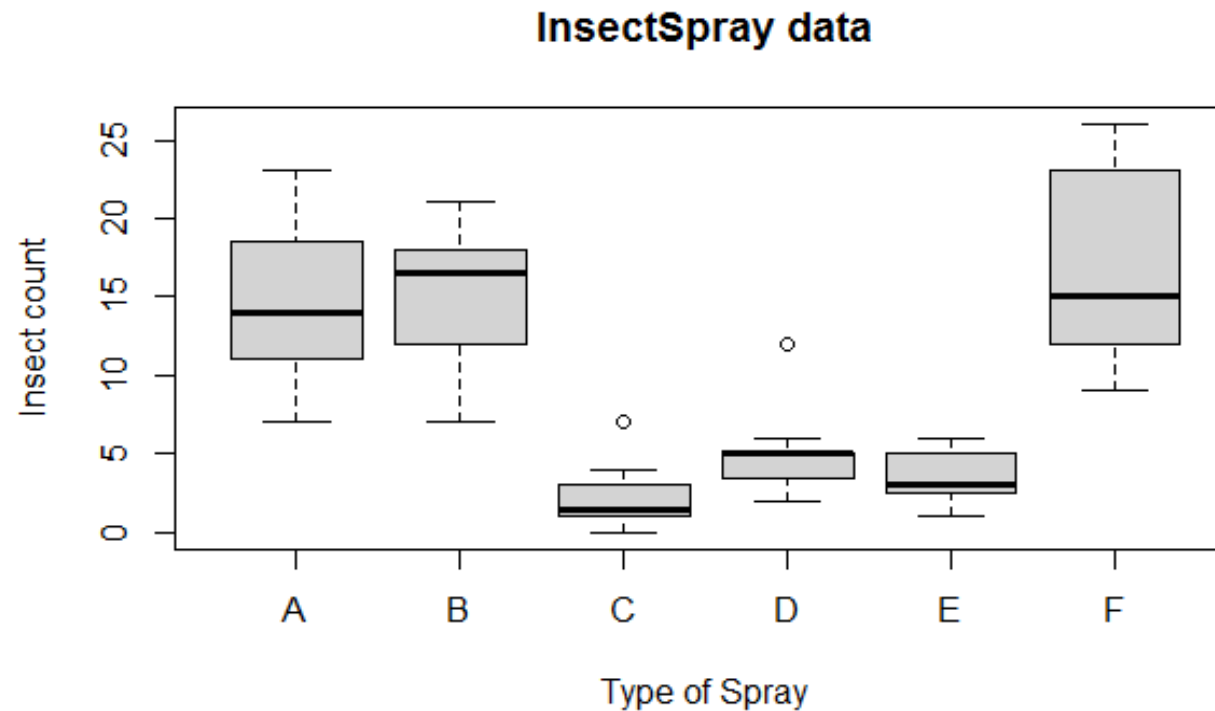
Data berikut menggunakan dataset yang disediakan dalam R. Dataset *InsectSprays* berisi jumlah serangga mati dalam desain percobaan pertanian yang terkena dampak akibat uji coba 6 macam insektisida (McNeil, D.(1977) *Interactive Data Analysis*. New York:Wiley). Pada contoh data ini, akan dilakukan uji perbedaan rata-rata serangga yang mati pada setiap kelompok insektisida yang digunakan. Dataset InsectSpray bertipe *data frame* yang terdiri dari 2 variabel (kolom) dengan masing-masing terdiri 72 observasi (baris).



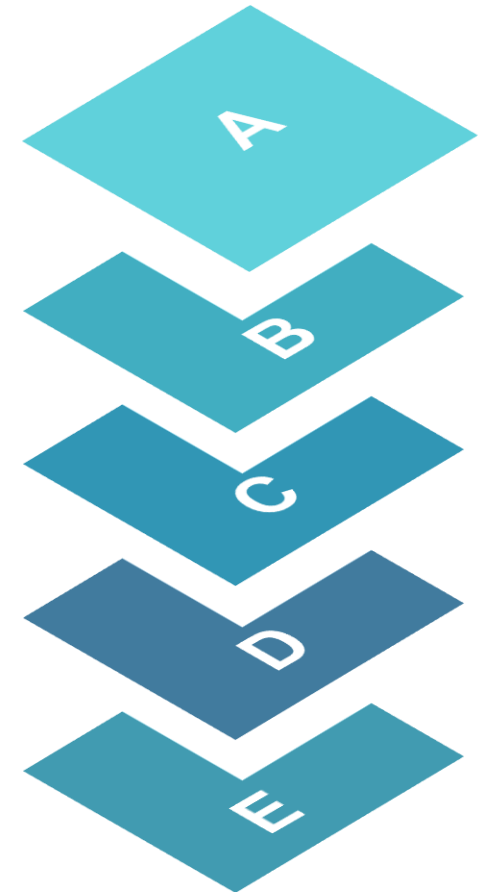
Langkah awal dilakukan dengan melakukan eskplorasi grafik dari data untuk mendapatkan gambaran visual dari data. Dengan menggunakan grafik box plot peneliti mendapatkan kesimpulan awal terkait dalam kelompok yang ada dari data tersebut.

Report					
Count					
Spray	Mean	Median	Maximum	Grouped Median	Minimum
A	14.91	14.00	23	14.00	7
B	15.33	16.50	21	16.25	7
C	2.08	1.50	7	1.67	0
D	4.92	5.00	12	4.57	2
E	3.50	3.00	6	3.40	1
F	16.67	15.00	26	15.00	9
Total	9.49	7.00	26	7.00	0

```
require(stats); require(graphics)
➤ boxplot(count ~ spray, data = Insect,
➤ +       xlab= "Type of Spray", ylab = "Insect count",
➤ +       main = "InsectSpray data", varwidth =TRUE,
➤ +       col = "lightgray")
```



Dari *boxplot* yang terbentuk, peneliti mampu melihat secara visual bahwa terdapat perbedaan dalam setiap grup. Untuk lebih meyakinkan peneliti maka akan dilakukan uji anova satu arah. Hipotesis awal dinyatakan bahwa tidak terdapat perbedaan kemampuan insektisida yang digunakan terhadap kematian serangga pada semua kelompok insektisida yang digunakan. Dengan tingkat signifikansinya sebesar 5%.



Sebelum dilakukan uji anova dilakukan pengujian asumsi terhadap kenormalan dan homogenitas varian dari datanya. Uji yang digunakan adalah uji normalitas menggunakan Uji Shapiro-Wilk test dan uji homogenitas menggunakan uji bartlet. Uji ini akan di analisis menggunakan R dan SPSS. Berikut seperti pada tampilan:

- by (data=count, INDICES = spray, FUN=shapiro.test)
- spray: A
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.95757, p-value = 0.7487
- -----
- spray: B
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.95031, p-value = 0.6415
- -----
- spray: C
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.85907, p-value = 0.04759
- -----

- spray: D
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.75063, p-value = 0.002713
- -----
- spray: E
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.92128, p-value = 0.2967
- -----
- spray: F
- Shapiro-Wilk normality test
- data: dd[x,]
- W = 0.88475, p-value = 0.1009
- -----
- bartlett.test(count~spray)
- Bartlett test of homogeneity of variances
- data: count by spray
- Bartlett's K-squared = 25.96, df = 5, p-value = 9.085e-05

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
A	.209	12	.156	.958	12	.749
a. Lilliefors Significance Correction						

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
B	.152	12	.200*	.950	12	.641
a. Lilliefors Significance Correction						
*. This is a lower bound of the true significance.						

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
C	.208	12	.158	.859	12	.048
a. Lilliefors Significance Correction						

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
D	.320	12	.001	.751	12	.003
a. Lilliefors Significance Correction						

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
E	.164	11	.200*	.922	11	.332
a. Lilliefors Significance Correction						
*. This is a lower bound of the true significance.						

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
F	.209	12	.153	.885	12	.101
a. Lilliefors Significance Correction						

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.500
Bartlett's Test of Sphericity	Approx. Chi-Square	17.098
	df	1
	Sig.	.000

Berdasarkan output bisa disimpulkan bahwa variabel setiap grup yang akan peneliti analisis, sebagian besar telah berdistribusi normal (Shapiro.test $p\text{-value} > 0.05$). Sedangkan varian antar grup dalam variabelnya memiliki perbedaan varian (barlett.test $p\text{-value} = 9.085e-05$). Beberapa asumsi kenormalan dan kehomogenan data terlanggar, hal ini dapat mengganggu kesimpulan akhir dari analisis of varian. Sebaiknya peneliti minimal melakukan eliminasi terhadap data outlier, untuk memastikan agar asumsi terpenuhi. Pembahasan eliminasi outlier ini akan dibahas lebih mendalam pada analisis statistik lanjutan.



```
> aov.Insect <- aov(count ~ spray, data = Insect)
> summary(aov.Insect)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
spray	5	2669	533.8	34.7	<2e-16 ***
Residuals	66	1015	15.4		

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

Hasil output dari one way anova dapat dilihat dari `summary(aov.Insect)`, ditampilkan output table anova. Analisis berdasarkan output tersebut dapat dilihat dari nilai p-value ($<2e-16$) dari uji F sangat signifikan, sehingga hipotesis awal ditolak. Kemudian dapat disimpulkan bahwa cukup bukti untuk menyatakan bahwa terdapat perbedaan efek insektisida terhadap kematian serangga. Karena uji anova ditolak hipotesis nol, maka langkah selanjutnya dilakukan uji perbandingan ganda (*multiple comparison*). Uji perbandingan berganda peneliti menggunakan uji Tukey HSD agar lebih terlihat pasangan kelompok mana yang berbeda signifikan secara statistic dan yang tidak signifikan berbeda.

```

> TukeyHSD(aov.Insect)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = count ~ spray, data = Insect)

$spray
      diff      lwr      upr    p adj
B-A  0.8333333 -3.866075  5.532742 0.9951810
C-A -12.4166667 -17.116075 -7.717258 0.0000000
D-A  -9.5833333 -14.282742 -4.883925 0.0000014
E-A -11.0000000 -15.699409 -6.300591 0.0000000
F-A   2.1666667 -2.532742  6.866075 0.7542147
C-B -13.2500000 -17.949409 -8.550591 0.0000000
D-B -10.4166667 -15.116075 -5.717258 0.0000002
E-B -11.8333333 -16.532742 -7.133925 0.0000000
F-B   1.3333333 -3.366075  6.032742 0.9603075
D-C   2.8333333 -1.866075  7.532742 0.4920707
E-C   1.4166667 -3.282742  6.116075 0.9488669
F-C  14.5833333  9.883925 19.282742 0.0000000
E-D  -1.4166667 -6.116075  3.282742 0.9488669
F-D  11.7500000  7.050591 16.449409 0.0000000
F-E  13.1666667  8.467258 17.866075 0.0000000

```

Hasil perbandingan kelompok insektida (spray) dapat dijelaskan bahwa yang tidak terdapat perbedaan secara statistik ($p\text{-value} > 5\%$) adalah antara pasangan kelompok insektisida jenis B dengan A, F dengan A, F dengan B, D dengan C, E dengan C dan pasangan antara E dan D. Selain hasil perbandingan antara pasangan kelompok tersebut maka pasangan tersebut adalah berbeda secara signifikan dalam mempengaruhi kematian serangga.



Analisis Pengaruh Dosis Vitamin C Terhadap Pertumbuhan Gigi Pada Hewan Marmut

Aplikasi : R Studio

Kasus berikut menggunakan dataset *ToothGrowth* dalam *package datasets*. Data ini merupakan hasil percobaan untuk melihat pengaruh/efek dari beberapa dosis vitamin c terhadap pertumbuhan gigi pada hewan marmut dengan menggunakan dua metode pemberian yang berbeda. Tiga dosis vitamin c yang diberikan antara lain : 0.5 mg/hari, 1 mg/hari dan 2 mg/hari. Sedangkan dua jenis metode pemberian vitamin C adalah menggunakan jus jeruk (OJ) dan dalam bentuk pil vitamin C (VC).



➤ # Explore

```
> data("ToothGrowth")
```

```
> Tooth <- ToothGrowth
```

```
> str(Tooth)
```

```
'data.frame':  60 obs. of  3 variables:
```

```
$ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
```

```
$ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
```

```
$ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Hasil output terlihat bahwa data frame Tooth memiliki 60 baris observasi dan 3 kolom variable (len (bertipe numerik/kontinu), sup (bertipe faktor/kategori) dan dose (bertipe num/kontinu). Karena di ingin mengelompokkan berdasarkan jumlah dosisnya, maka variabel dose perlu peneliti rubah menjadi bertipe kategori/factor menggunakan perintah `Tooth$dose<-as.factor(Tooth$dose)` seperti di bawah ini.

```

> Tooth$dose<-as.factor(Tooth$dose)
> str(Tooth)
'data.frame':  60 obs. of  3 variables:
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
 $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
➤ bartlett.test(len ~supp, data=Tooth)

```

Bartlett test of homogeneity of variances

data: len by supp

Bartlett's K-squared = 1.4217, df = 1, p-value = 0.2331

```

> bartlett.test(len ~dose, data=Tooth

```

Bartlett test of homogeneity of variances

data: len by dose

Bartlett's K-squared = 0.66547, df = 2, p-value = 0.717

Uji yang digunakan menggunakan uji Bartlett, terlihat nilai p-value dari uji pada kedua kelompok lebih besar dari standar tingkat signifikansi 5%. Maka peneliti menerima hipotesis nol dan menyatakan bahwa variansi dari kelompok ada homogen.

```

> #Anova Two Ways
> aov.vitc<-aov(len ~ supp + dose, data=Tooth)
> summary(aov.vitc)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.4	205.4	14.02	0.000429 ***
dose	2	2426.4	1213.2	82.81	< 2e-16 ***
Residuals	56	820.4	14.7		

```

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

```

Hasil output terlihat pada bagian p-value untuk variable supp dan dose memiliki tingkat signifikansi yang tinggi. Yang berarti pengaruh kedua variabel tersebut dalam percobaan panjang pertumbuhan gigi marmut signifikan. Setiap kelompok percobaan dalam variable tersebut memberikan pertumbuhan rata-rata panjang gigi marmut yang tidak sama. Langkah berikutnya adalah melakukan perbandingan berganda untuk mencari kelompok mana dari kedua variabel tersebut yang sangat berbeda secara rata-rata pertumbuhannya menggunakan fungsi TukeyHSD.

```
> TukeyHSD(aov.vitc)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = len ~ supp + dose, data = Tooth)

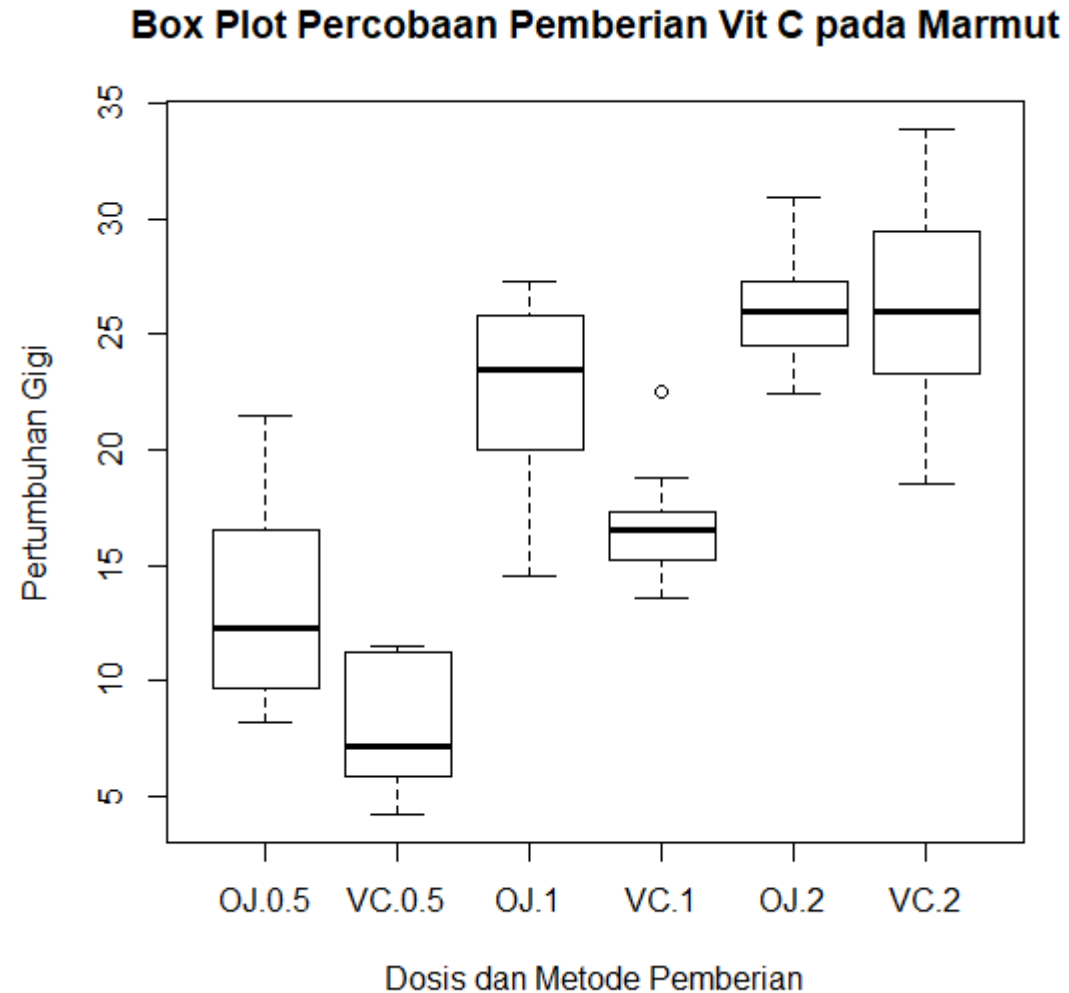
$supp
      diff      lwr      upr    p adj
VC-0J -3.7 -5.679762 -1.720238 0.0004293

$dose
      diff      lwr      upr p adj
1-0.5  9.130  6.215909 12.044091 0e+00
2-0.5 15.495 12.580909 18.409091 0e+00
2-1    6.365  3.450909  9.279091 7e-06
```

Hasil output terlihat pada bagian p-value untuk perbandingan kelompok dalam variabel supp memberikan hasil pertumbuhan yang signifikan berbeda secara rata-rata antara menggunakan jus jeruk dan pemberian menggunakan vitamin C. Pada variabel dose, percobaan pemberian dosis vitamin C memberikan pengaruh rata-rata pertumbuhan panjang gigi marmut yang berbeda secara signifikan pada dosis 1 mg/day dengan 0.5 mg/day serta pada kelompok dosis 2 mg/day dengan 0.5 mg/day. Sedangkan percobaan pemberian dosis antara 2 mg/day dengan 1 mg/day tidak memberikan rata-rata pertumbuhan yang signifikan berbeda secara statistic. Untuk lebih menjelaskan hasil yang didapatkan, peneliti dapat melihat secara visual menggunakan *box plot*.

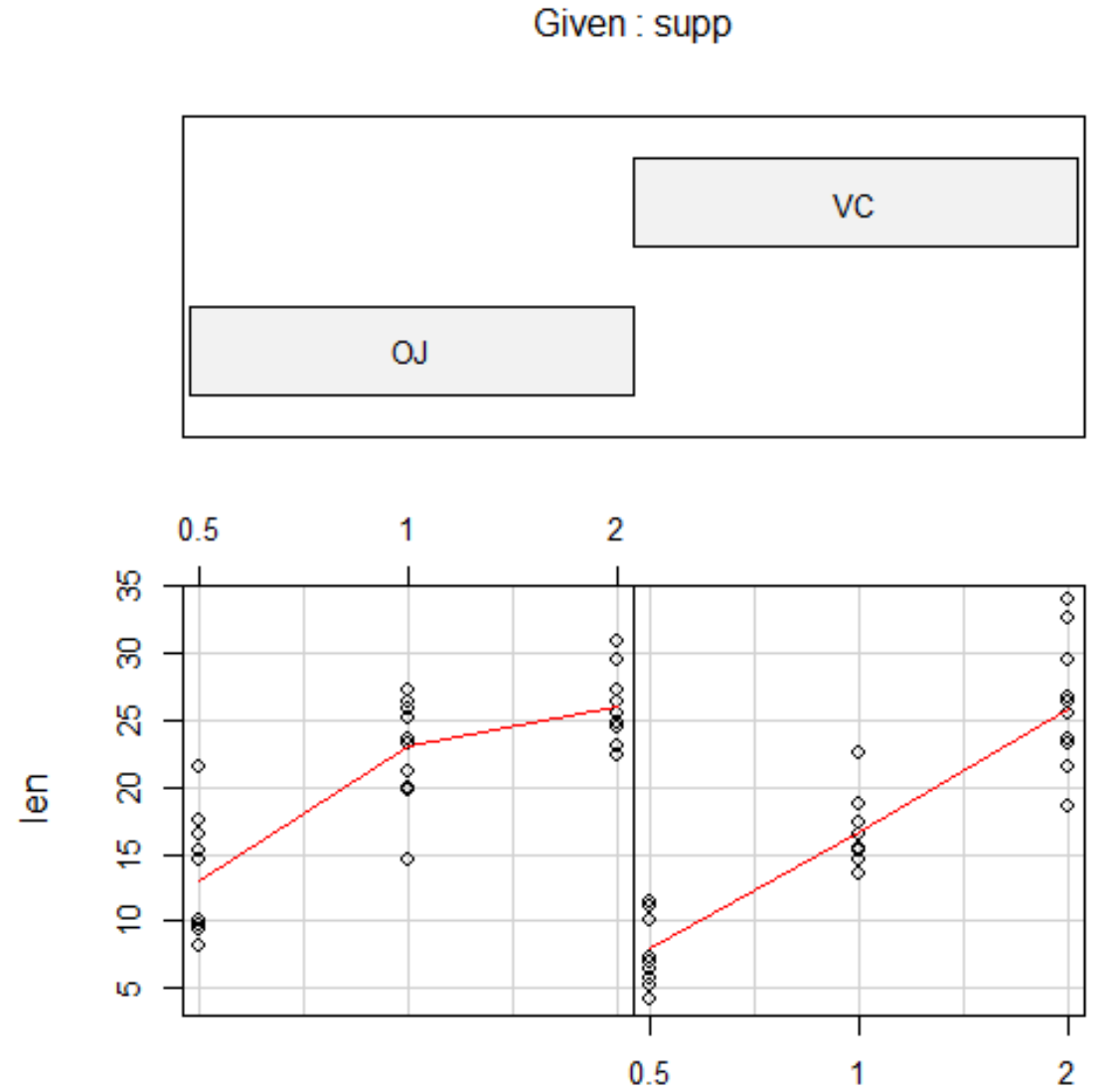


```
> boxplot(len ~ supp + dose, data=Tooth, xlab="Dosis dan Metode Pemberian",  
ylab="Pertumbuhan Gigi", main="Box Plot Percobaan Pemberian Vit C pada Marmut")
```



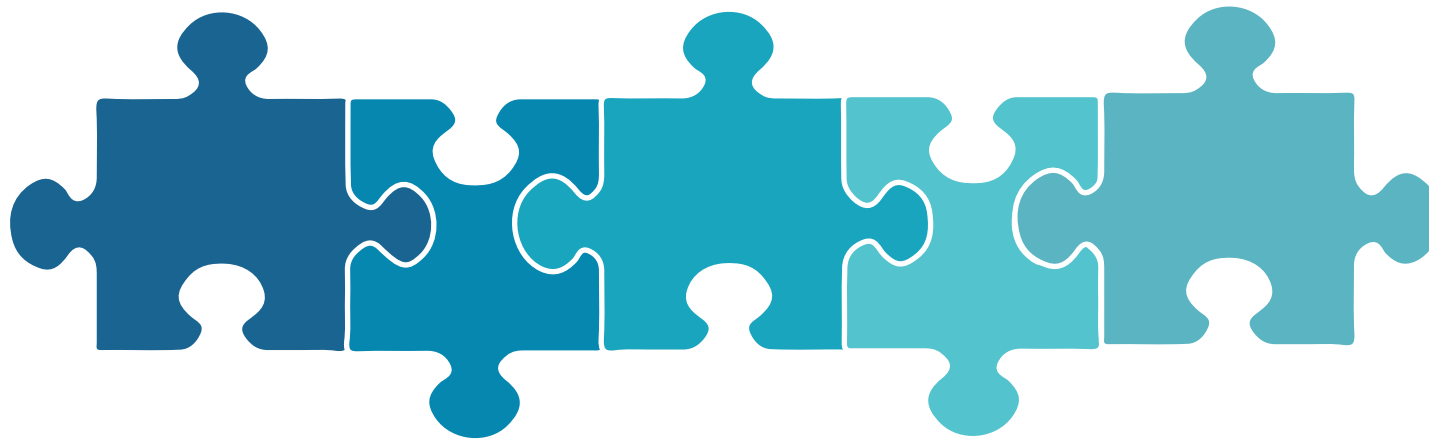
Peneliti juga dapat membuat visual hasil percobaan ini menggunakan conditioning plot menggunakan fungsi `coplot`. Fungsi ini akan menampilkan keragaman yang diberikan pada setiap kelompok variable percobaan.

```
oplot(len ~ dose | supp, data = Tooth, panel  
= panel.smooth,  
+ xlab = "ToothGrowth data: length vs dose,  
given type of supplement")
```



ToothGrowth data: length vs dose, given type of supplement

Berdasarkan hasil dari *boxplot* dan *conditioning plot*, terlihat posisi panjang pertumbuhan gigi marmut berdasarkan metode pemberian vitamin C memberikan pengaruh yang berbeda. Pertumbuhan gigi marmut dengan menggunakan metode pemberian menggunakan jus jeruk memberikan pengaruh lebih tinggi dibandingkan dengan vitamin C. Hasil tersebut berbanding lurus dengan hasil uji anova dan perbandingan berganda sebelumnya. Begitu pula untuk perbandingan dosis pemberian vitamin C dengan berbagai dosisnya memberikan hasil yang serupa dengan hasil pada uji anova dan perbandingan berganda sebelumnya.





Simulasi Pendanaan Program Pensiun Manfaat Pasti dengan Metode *Spreading Gains and Losses*

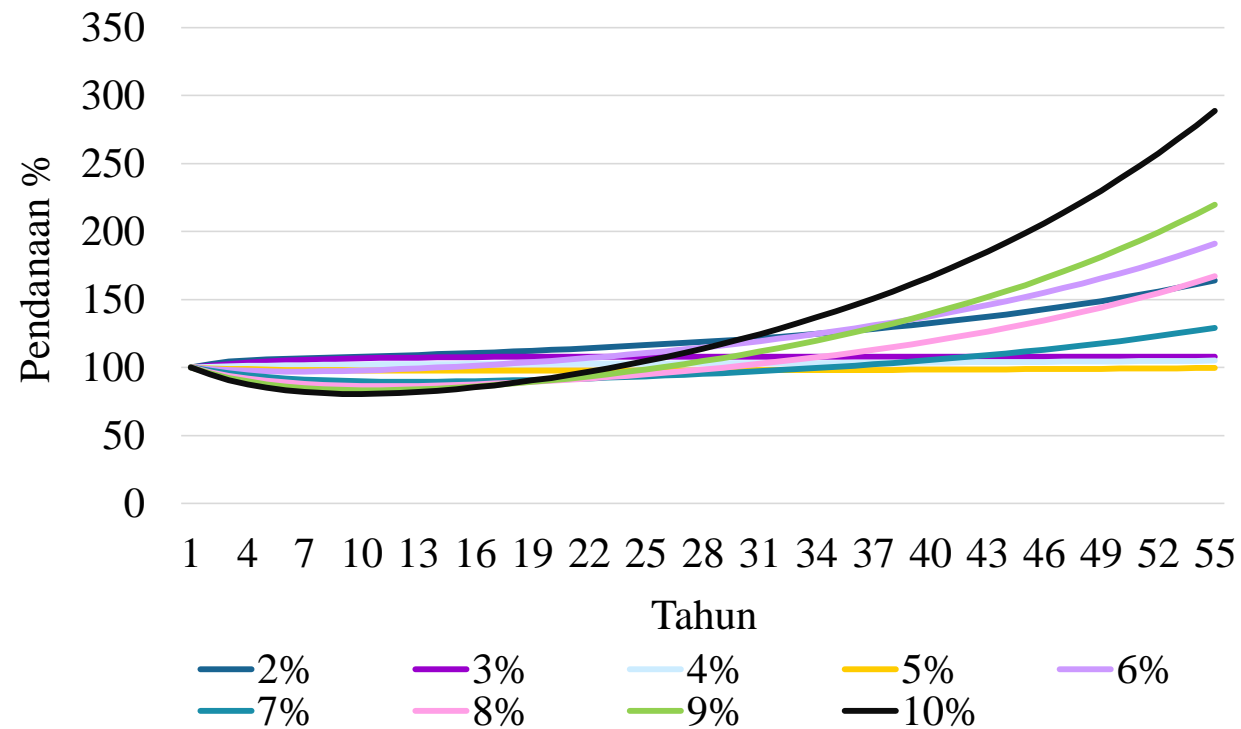
Aplikasi : Microsoft Excel

Penerapan metode *Spreading Gains and Losses* pada pendanaan program pensiun manfaat pasti. Data peluang hidup laki-laki diperoleh menggunakan Tabel Mortalita Indonesia (TMI) tahun 2019. Studi kasus Pegawai Negeri Sipil, mengikuti pendanaan program pensiun manfaat pasti dengan usia yang berbeda-beda yaitu sebesar 20 tahun hingga 35 tahun. batas usia pensiun adalah 56 tahun ($r = 56$).



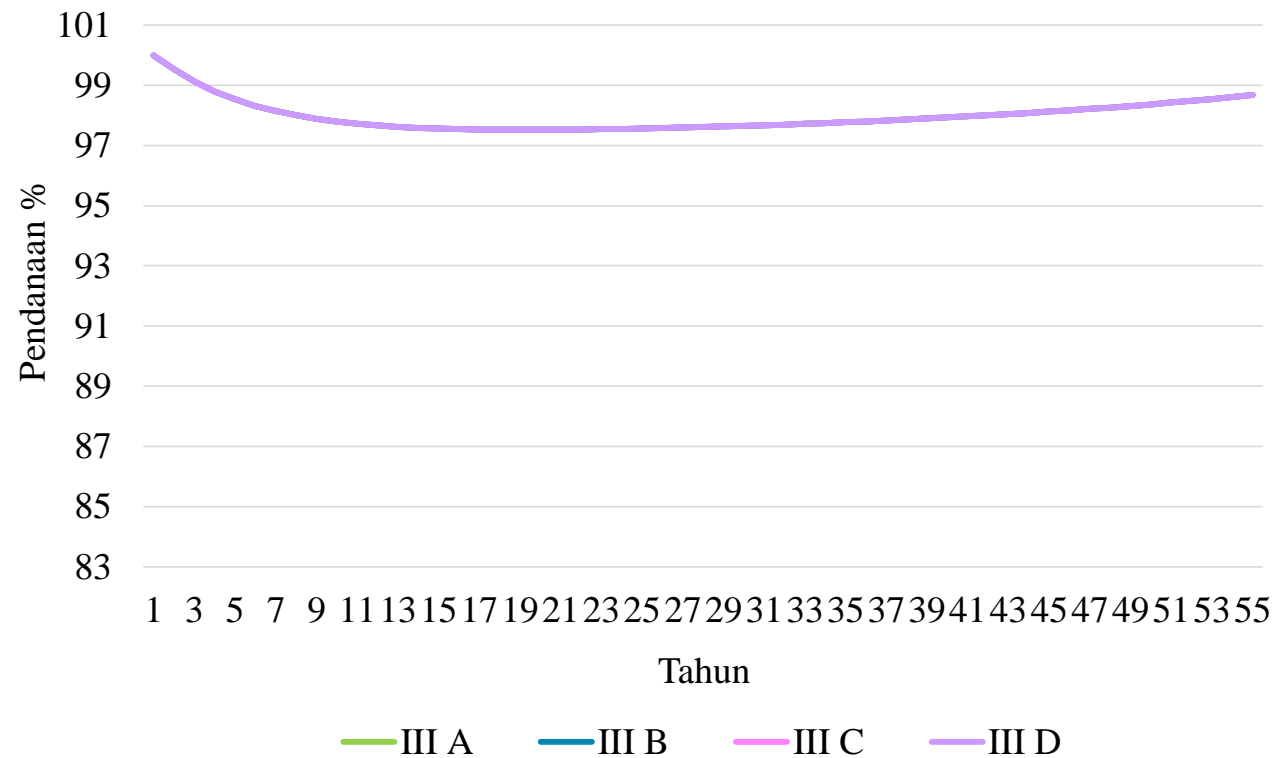
**Persentase Pendanaan Program Pensiun Manfaat Pasti dengan Asumsi Tingkat Suku Bunga Pengembalian
Investasi Aktuaria Berbeda-beda**

Pendanaan program pensiun manfaat pasti terdiri dari persentase total dana yang dimiliki suatu perusahaan terhadap dana yang harus dicadangkan. Hasil dari pendanaan program pensiun manfaat pasti dengan gaji sebesar Rp2.688.500, usia peserta sebesar 20 tahun dan asumsi tingkat suku bunga pengembalian investasi aktuarial (i_A) yang berbeda-beda dapat dilihat sebagai berikut:



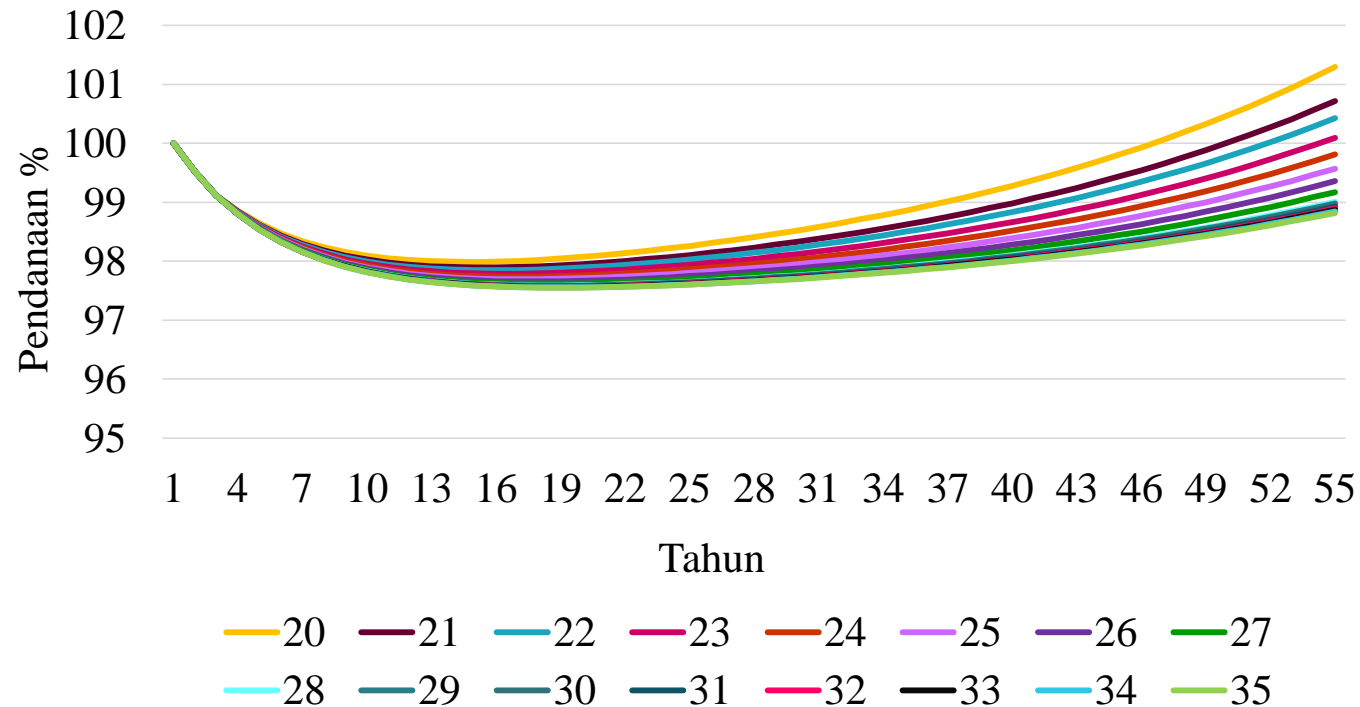
Persentase Pendanaan Program Pensiun Manfaat Pasti dengan Gaji Pegawai Negeri Sipil Berbeda-beda

Hasil dari pendanaan program pensiun manfaat pasti dengan usia peserta yang sama yaitu sebesar 20 tahun, asumsi tingkat suku bunga pengembalian investasi aktuarial (i_A) yang sama yaitu sebesar 5% dan peserta yang berbeda-beda yaitu:



Persentase Pendanaan Program Pensiun Manfaat Pasti dengan Usia Pegawai Negeri Sipil Berbeda-beda

Hasil dari pendanaan program pensiun manfaat pasti dengan gaji peserta yang bersifat konstan yaitu sebesar Rp2.688.500. Asumsi tingkat suku bunga pengembalian investasi aktuarial (i_A) yang digunakan sama yaitu sebesar 5%. Usia peserta yang berbeda-beda yaitu:



Berdasarkan dari hasil penelitian yang telah dibahas, maka dapat disimpulkan bahwa perbedaan asumsi tingkat suku bunga pengembalian investasi aktuarial (i_A) yang berbeda-beda, dapat mempengaruhi pendanaan program pensiun manfaat pasti dalam jangka panjang. Perbedaan asumsi tingkat suku bunga pengembalian investasi aktuarial (i_A) dengan tingkat suku bunga aktual (i) dapat menyebabkan kesenjangan pada pendanaan program pensiun manfaat pasti. Jika dilihat dari perilaku jangka panjang, menggunakan asumsi tingkat suku bunga pengembalian investasi aktuarial yang optimal dapat menstabilkan pendanaan program pensiun manfaat pasti. Saat asumsi tingkat suku bunga pengembalian investasi aktuarial sebesar 5%, maka pendanaan program pensiun manfaat pasti akan stabil di tahun ke-8 yaitu sebesar 0,109% dan dalam jangka panjang pendanaan program pensiun manfaat pasti bernilai stabil.

Pada saat pendanaan program pensiun manfaat pasti menggunakan gaji Pegawai Negeri Sipil yang berbeda-beda. Besarnya pendanaan program pensiun manfaat pasti bernilai sama. Maka menggunakan gaji peserta yang berbeda-beda tidak mempengaruhi pendanaan program pensiun manfaat pasti dalam jangka panjang. Pada saat pendanaan program pensiun manfaat pasti menggunakan usia masuk Pegawai Negeri Sipil yang berbeda-beda dapat mempengaruhi pendanaan program pensiun manfaat pasti. Jika dilihat dari perilaku jangka panjang maka semakin kecil usia peserta yang mengikuti pendanaan program pensiun manfaat pasti maka semakin cepat kerugian pada pendanaan dapat tertutupi. Hal ini dikarenakan laju kematian di usia 20 tahun lebih rendah sehingga pendanaan program pensiun manfaat pasti mengalami keuntungan. Pendanaan program pensiun manfaat pasti dengan menggunakan usia yang berbeda-beda dapat mempengaruhi pendanaan program pensiun manfaat pasti dalam jangka panjang.



Pengaruh Konsumsi Daging Sapi dan Daging Ayam Terhadap Tingkat Permintaan Daging Perkapita di Amerika Serikat

Aplikasi: SPSS

Pengaruh Konsumsi Daging Sapi dan Daging Ayam Terhadap Tingkat Permintaan Daging Perkapita di Amerika Serikat Selama Periode 1960-1982. Berikut disajikan data kasus permintaan daging di Amerika Serikat pada tahun 1960-1982:

Tahun	Tingkat Permintaan Daging Perkapita	Konsumsi Daging Sapi	Konsumsi Daging Ayam
1960	27.8	42.2	50.7
1961	29.9	38.1	52
1962	29.8	40.3	54
1963	30.8	39.5	55.3
1964	31.2	37.7	54.7
1965	33.3	38.1	63.7
1966	35.6	39.3	69.8
1967	36.4	37.8	65.9
1968	36.7	38.4	64.5
1969	38.4	40.1	70
1970	40.4	38.6	73.2
1971	40.3	39.8	67.8
1972	41.8	39.7	79.1
1973	40.4	52.1	95.4
1974	40.7	48.9	94.2
1975	40.1	58.3	123.5
1976	42.7	57.9	129.9
1977	44.1	56.5	117.6
1978	46.7	63.7	130.9
1979	50.6	61.6	129.8
1980	50.1	58.9	128
1981	51.7	66.4	141
1982	52.9	70.4	168.2

Satuan: ton

Hasil Uji F dan Uji T

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1037.577	2	518.789	65.524	.000 ^a
	Residual	158.352	20	7.918		
	Total	1195.929	22			

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	32.835	4.818		6.815	.000
	X1	-.512	.222	-.770	-2.304	.032
	X2	.347	.070	1.660	4.962	.000

a. Dependent Variable: Y

Hasil R *square*

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.931 ^a	.868	.854	2.81382	.868	65.524	2	20	.000

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Hasil R *square* adalah 0.868 % atau 86%

Model Regresi

$$Y = 32.835 - 0,512X1 + 0,347X2$$

Uji Normalitas

Uji Normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel, apakah sebaran data tersebut berdistribusi normal ataukah tidak.

Hasil Uji Normalitas

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Unstandardized Residual	23	-5.80422	4.27871	-9.16E-15	2.68287134	-.394	.481	.220	.935
Valid N (listwise)	23								

Berdasarkan hasil dari *Descriptive Statistics* dapat dihitung rasio skewness = $-0,394/0,481 = -0,819$ sementara nilai rasio kurtosis $0,220/0.935 = 0,235$. karena rasio skewness dan rasio kurtosis berada di rentang -2 hingga 2. Maka data tersebut berdistribusi normal.

Uji Multikolinearitas

Situasi dimana terdapat korelasi atau hubungan linear antar variabel-variabel bebas diantara satu dengan yang lainnya

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	32.835	4.818		6.815	.000		
	X1	-.512	.222	-.770	-2.304	.032	.059	16.895
	X2	.347	.070	1.660	4.962	.000	.059	16.895

a. Dependent Variable: Y

Hasil Uji Multikolinearitas

Berdasarkan hasil output SPSS pada table “*Coeficients*” menunjukkan bahwa variable Independen memiliki nilai Tolerance kurang dari 0,10. Hasil perhitungan nilai Variance Inflation Factor (VIF) juga menunjukkan hal yang sama semua variable independen yang memiliki nilai VIF lebih besar dari 10. Jadi dapat disimpulkan bahwa terjadi multikolinearitas antar variable Independen dalam model regresi.

Uji Autokorelasi?

Uji autokorelasi digunakan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik. Autokorelasi adalah korelasi yang terjadi antara residual pada satu pengamatan dengan pengamatan lain pada model regresi.

Hasil Uji Autokorelasi

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.931 ^a	.868	.854	2.81382	.838

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Nilai Durbin-Watson sebesar 0,838. Nilai ini akan dibandingkan dengan nilai DW tabel. Di antara 0 dan nilai dL sebesar 1,1682. Jadi nilai DW lebih kecil dari nilai dL. Sehingga dapat disimpulkan model ini mengalami gejala autokorelasi positif.

Salah satu cara mengatasi autokorelasi yaitu dengan transformasi lag

Hasil Transformasi Uji Autokorelasi

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.652 ^a	.425	.361	1.71382	1.531
a. Predictors: (Constant), Lag_Lag_X2, Lag_Lag_X1					
b. Dependent Variable: Lag_Lag_Y1					

Setelah data di transformasi. Nilai Durbin-Watson menjadi 1.531. Nilai ini akan dibandingkan dengan nilai DW tabel. Di antara 0 dan nilai dL sebesar 1,1682. Jadi nilai DW lebih besar dari nilai dL. Sehingga dapat disimpulkan model ini bebas dari autokorelasi. Sehingga model regresinya menjadi :

$$Y=9.183-0.205X1+0.155X2$$

Uji Heterokedastisitas

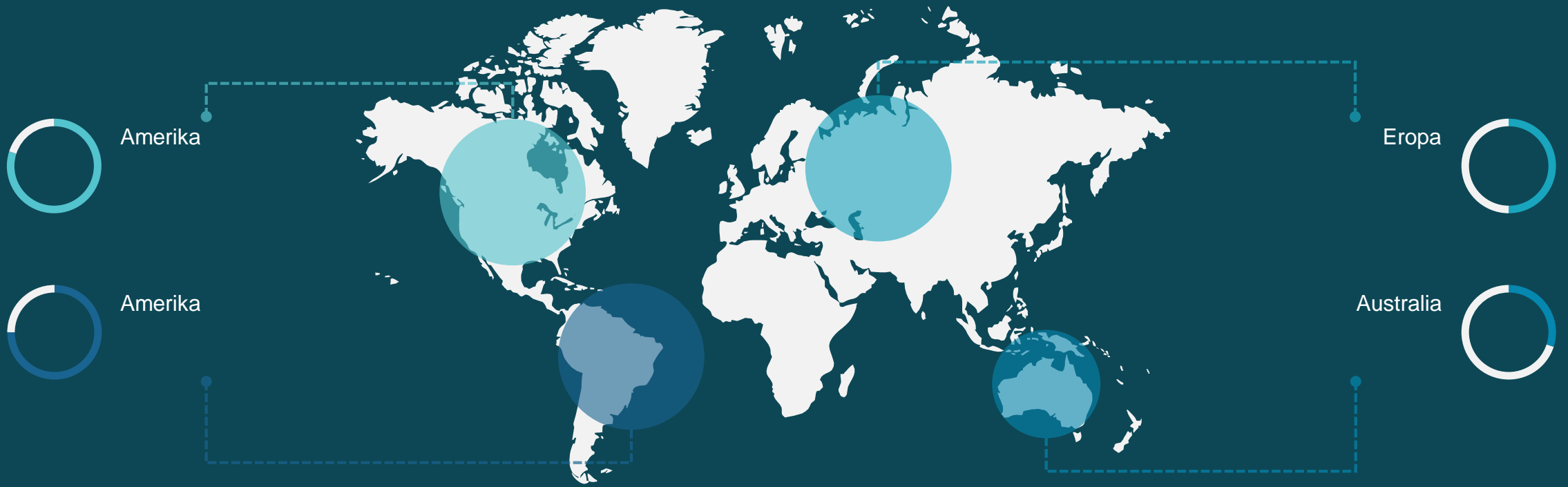
Uji heterokedastisitas berarti situasi dimana keragaman variable independen bervariasi pada data yang di miliki

Hasil Uji Heterokedastisitas

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	
1	(Constant)	1.199	2.469		.486
	X1	-.057	.114	-.396	.623
	X2	.040	.036	.876	1.107

a. Dependent Variable: Abs_RES

Pada hasil output dari SPSS menunjukkan bahwa tidak ada satupun konsumsi daging sapi dan konsumsi daging ayam yang signifikan secara statistik mempengaruhi permintaan daging perkapita. Hal ini terlihat dari hasil signifikan yang semuanya lebih besar dari $\alpha=0,05$. Maka model regresi ini tidak mengandung heterokedastisitas atau bersifat homoskedastisitas.



Pada uji asumsi klasik dari data konsumsi daging sapi dan daging ayam terhadap tingkat permintaan daging perkapita di Amerika Serikat. Data tersebut mengalami autokorelasi. Yang mana nilai DW autokorelasi sebesar 0,838. Setelah itu dilakukan transformasi sebanyak dua kali dengan nilai DW sebesar 1,531 sehingga tidak mengalami autokorelasi lagi. Sehingga dapat disimpulkan bahwa konsumsi daging sapi dan konsumsi daging ayam berpengaruh terhadap permintaan daging perkapita di Amerika Serikat pada periode 1960-1982.



Regresi Linear Pada Data Deret Waktu Kasus Kejahatan yang terjadi di Amerika Serikat periode tahun 1997–2016.

Aplikasi: KNIME

Data yang di analisis adalah data milik FBI Amerika Serikat yang sudah dipublikasikan. Data tersebut merupakan rekapitulasi pelanggaran hukum (kejahatan) yang terjadi di Amerika Serikat pada periode tahun 1997–2016.

Table 1																
Crime in the United States																
by Volume and Rate per 100,000 Inhabitants, 1997–2016																
Year	Population ¹	Violent crime ²	Violent crime rate	Murder and nonnegligent manslaughter	Murder and nonnegligent manslaughter rate	Rape (revised definition ³)	Rape (revised definition) rate ³	Rape (legacy definition ⁴)	Rape (legacy definition) rate ⁴	Robbery	Robbery rate	Aggravated assault	Aggravated assault rate	Property crime	Property crime rate	Burglary
1997	267,783,607	1,636,096	611.0	18,208	6.8			96,153	35.9	498,534	186.2	1,023,201	382.1	11,558,475	4,316.3	2,460,526
1998	270,248,003	1,533,887	567.6	16,974	6.3			93,144	34.5	447,186	165.5	976,583	361.4	10,951,827	4,052.5	2,332,735
1999	272,690,813	1,426,044	523.0	15,522	5.7			89,411	32.8	409,371	150.1	911,740	334.3	10,208,334	3,743.6	2,100,739
2000	281,421,906	1,425,486	506.5	15,586	5.5			90,178	32.0	408,016	145.0	911,706	324.0	10,182,584	3,618.3	2,050,992
2001 ⁵	285,317,559	1,439,480	504.5	16,037	5.6			90,863	31.8	423,557	148.5	909,023	318.6	10,437,189	3,658.1	2,116,531
2002	287,973,924	1,423,677	494.4	16,229	5.6			95,235	33.1	420,806	146.1	891,407	309.5	10,455,277	3,630.6	2,151,252
2003	290,788,976	1,383,676	475.8	16,528	5.7			93,883	32.3	414,235	142.5	859,030	295.4	10,442,862	3,591.2	2,154,834
2004	293,656,842	1,360,088	463.2	16,148	5.5			95,089	32.4	401,470	136.7	847,381	288.6	10,319,386	3,514.1	2,144,446
2005	296,507,061	1,390,745	469.0	16,740	5.6			94,347	31.8	417,438	140.8	862,220	290.8	10,174,754	3,431.5	2,155,448
2006	299,398,484	1,435,123	479.3	17,309	5.8			94,472	31.6	449,246	150.0	874,096	292.0	10,019,601	3,346.6	2,194,993
2007	301,621,157	1,422,970	471.8	17,128	5.7			92,160	30.6	447,324	148.3	866,358	287.2	9,882,212	3,276.4	2,190,198
2008	304,059,724	1,394,461	458.6	16,465	5.4			90,750	29.8	443,563	145.9	843,683	277.5	9,774,152	3,214.6	2,228,887
2009	307,006,550	1,325,896	431.9	15,399	5.0			89,241	29.1	408,742	133.1	812,514	264.7	9,337,060	3,041.3	2,203,313
2010	309,330,219	1,251,248	404.5	14,722	4.8			85,593	27.7	369,089	119.3	781,844	252.8	9,112,625	2,945.9	2,168,459

Pisahkan data yang tidak diperlukan (pra-proses)

Pada analisis hanya akan menggunakan tiga kolom untuk dianalisis: *Year* (tahun) , *Population* (populasi), dan *Violent crime rate* (kadar kejahatan kekerasan).

Year	Population	<i>Violent_crime_rate</i>
1997	267,783,607	611.0
1998	270,248,003	567.6
1999	272,690,813	523.0
2000	281,421,906	506.5
2001	285,317,559	504.5
2002	287,973,924	494.4
2003	290,788,976	475.8
2004	293,656,842	463.2
2005	296,507,061	469.0
2006	299,398,484	479.3
2007	301,621,157	471.8
2008	304,059,724	458.6
2009	307,006,550	431.9
2010	309,330,219	404.5
2011	311,587,816	387.1
2012	313,873,685	387.8
2013	316,497,531	369.1
2014	318,907,401	361.6
2015	320,896,618	373.7
2016	323,127,513	386.3

Normalisasi tahap pertama :

Terdapat data-data yang mungkin tidak sesuai dengan kebutuhan aplikasi atau algoritma data *science*.

Misalnya, pada tabel terdapat nilai sel 2015^x. Padahal *normalnya*, semua nilai sel-sel yang dibutuhkan adalah tanpa tanda pangkat x. Seharusnya nilai sel tersebut adalah 2015 saja. Ubah sel tersebut sama dengan dua cara.

Normalisasi tahap kedua :

Mengubah tipe data dari *string* menjadi *integer/number*. Untuk tahap normalisasi yang ini, digunakan aplikasi

Knime.

KNIME Analytics Platform

File Edit View Node Help

KNIME Explorer

- My-KNIME-Hub (api.hub.knime.com)
- EXAMPLES (knime@api.hub.knime.com)
- LOCAL (Local Workspace)
 - Example Workflows
 - FBI-regresi
 - KNIME_project

Node Repository

- Views
- Analytics
 - Mining
 - Bayes
 - Clustering
 - Rule Induction
 - Neural Network
 - Decision Tree
 - Decision Tree Ensemble
 - Misc Classifiers
 - Ensemble Learning
 - Item Sets / Association Rules
 - Linear Discriminant Analysis
 - Linear/Polynomial Regression
 - Linear Regression Learner
 - Polynomial Regression Learner
 - Regression Predictor
 - Logistic Regression
 - MDS
 - PCA
 - PMML
 - SVM

Workflow: Excel Reader (Node 1) → String To Number (Node 2) → Linear Regression Learner (Node 3)

String To Number (3:2)

State: EXECUTED

Port Output: Port 0 Load data Rows: 20, C

ID	Year	Population	Violent_crime_rate
Row0	1997.0	267783607	611.0
Row1	1998.0	270248003	567.6
Row2	1999.0	272690813	523.0
Row3	2000.0	281421906	506.5
Row4	2001.0	285317559	504.5
Row5	2002.0	287973924	494.4
Row6	2003.0	290788976	475.8
Row7	2004.0	293656842	463.2
Row8	2005.0	296507061	469.0
Row9	2006.0	299398484	479.3
Row10	2007.0	301621157	471.8
Row11	2008.0	304059724	458.6
Row12	2009.0	307006550	431.9
Row13	2010.0	309330219	404.5

String To Number

Converts strings in a column (or a set of columns) to numbers. If the node fails to parse a string, it will generate a missing cell and append a warning message to the KNIME Console with detailed information.

Dialog Options

Type

Choose the DataType that your string should be converted to.

Decimal Separator

Menemukan Model

Selanjutnya ditemukan persamaan yang cocok untuk deretan data. Saat ini menggunakan metode *linear regression*. Persamaan akhirnya adalah berpola $y = a + bx$

y = adalah kadar kejahatan kekerasan (*violent crime rate*).

x = tahun

a = koefisien titik potong sumbu y (ketika $x = 0$) atau disebut juga *intercept*

b = koefisien / konstanta kemiringan.

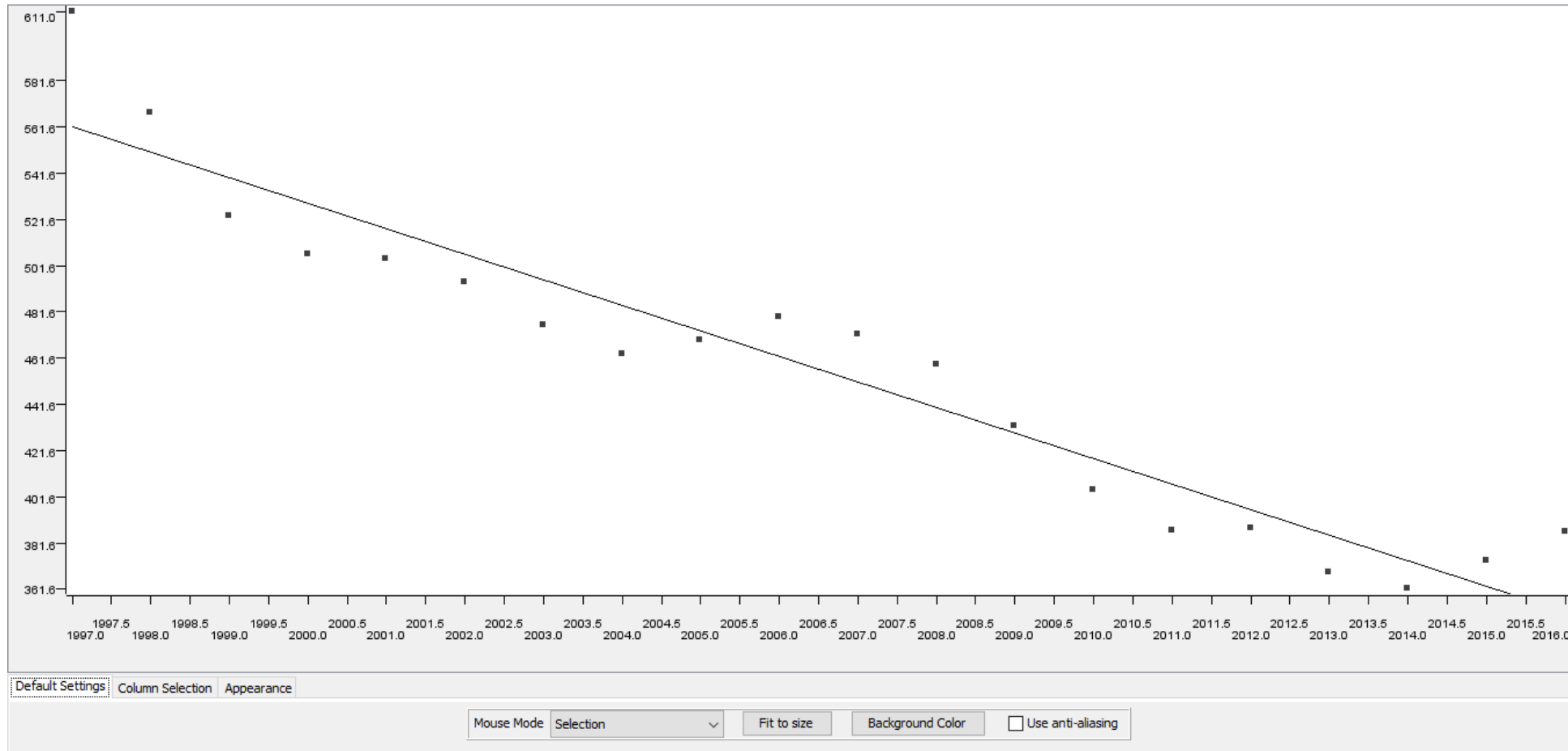
Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Year	-11.065	0.8207	-13.4827	7.57E-11
Intercept	22,658.3329	1,646.7125	13.7597	5.41E-11

Multiple R-Squared: 0.9099

Adjusted R-Squared: 0.9049

Melihat output hasil pemodelan



Dari output di atas, maka didapatkan:

koefisien a = 22.658,3329

koefisien b = -11,065

Model yang di dapatkan:

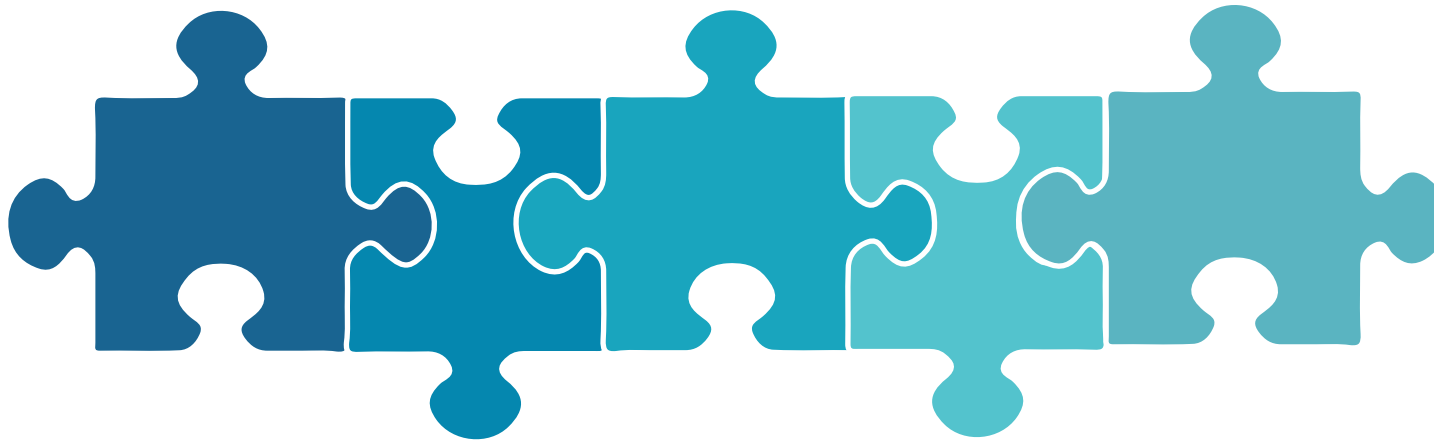
$$y = 22.658,3329 - 11,065x$$

Menguji Model

Nilai x yaitu tahun 2000 maka dapat diterapkan dalam model

$Y = 22.658,3329 - 11,065x$ (2000) = 528,3329 (kadar kejahatan kekerasan di tahun 2000).

Maka kasus kejahatan yang terjadi di Amerika Serikat pada tahun 2000 sebesar 528 kejahatan.



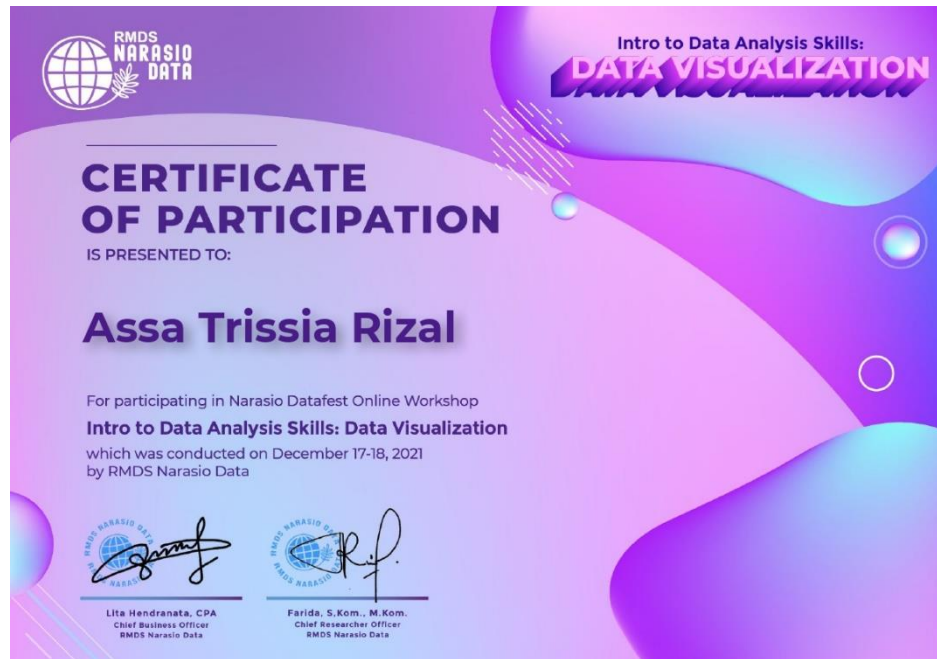


Analisis Regresi Linear Berganda

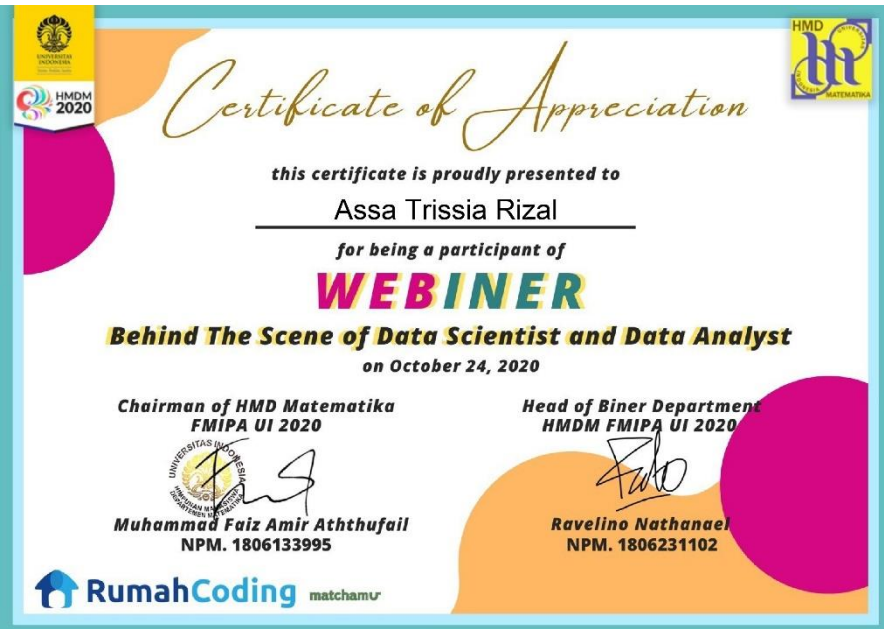
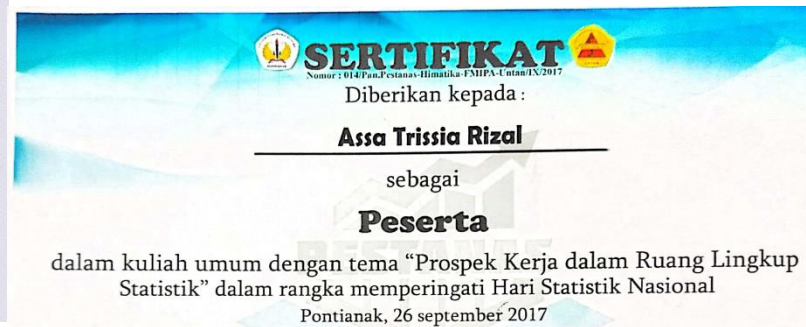
Aplikasi: Minitab

Dalam suatu kelurahan akan dilakukan penelitian terhadap 10 rumah tangga yang dipilih secara acak. Setelah dilakukan observasi, didapatkan data pendapatan per minggu (X_1), jumlah anggota keluarga (X_2), dan pengeluaran untuk pembelian barang-barang tahan lama per minggu (Y) sebagai berikut.

Sertifikat-Sertifikat Data Science



Sertifikat-Sertifikat Data Science



Sertifikat-Sertifikat Data Science



#DQLABSQLT1CNMOPR

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
Fundamental SQL Using SELECT Statement

Oct 22, 2021



#DQLABINTR10JKWQV

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
R Fundamental for Data Science

Oct 21, 2021



#DQLABBGINRANDEUR

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
Introduction to Data Science with R

Oct 18, 2021



#DQLABINTP1MOVVRJP

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
Python Fundamental for Data Science

Oct 22, 2021



#DQLABQBDA1LDMDQS

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
Quiz Basic Analytics

Oct 21, 2021



#DQLABINTR10JKWQV

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Assa Trissia Rizal

Has Completed in
R Fundamental for Data Science

Oct 21, 2021



TERIMA KASIH