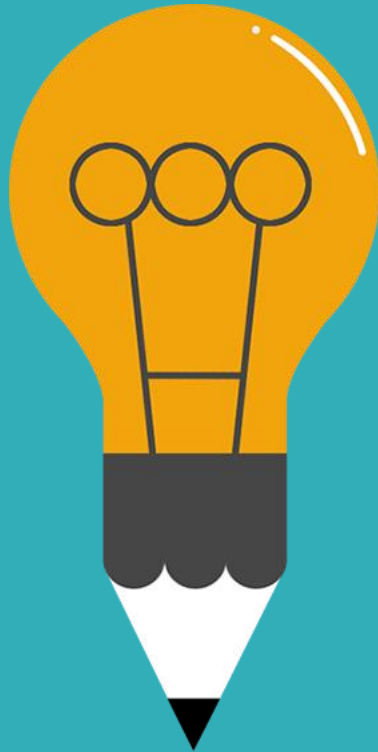


READI

Risk Management, Economic Sustainability
and Actuarial Science Development in Indonesia



PERSATUAN AKTUARIS INDONESIA
(THE SOCIETY OF ACTUARIES OF INDONESIA)



INVESTMENT GROUP

Mohamad Atok
Silvia R Dewi
Daud Kurniawan
Aulia Fahreza
Farhan Trunna
Milla Sejahtera

Bogor, 11-13 Februari 2019

Regression Analysis of Data Insurance using R



Agenda



01 Objective and Scope

02 Data

03 Method

04 R Program Syntax

05 Result

06 Summary



OBJECTIVE AND SCOPE

Objectives:

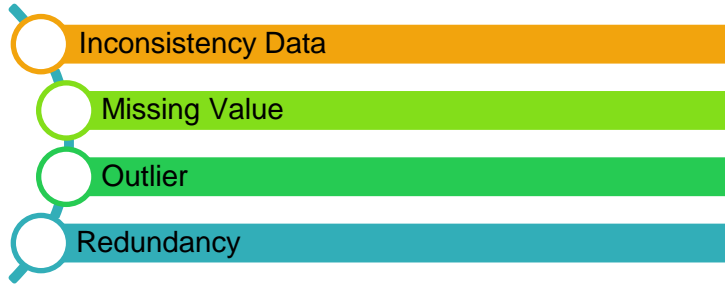
- Able to build model which only contain influencing variable toward the sum assured
- Able to predict sum assured based on the model

Scope:

- Individual line of business only

DATA

Pre-processing Data (Identifying and Cleansing)



Variable Data

Dependent Variable :
Uang Pertanggung
(Sum Assured)

Independent Variable :
Kelas Bisnis (business class)
Pekerjaan (Job)
Usia Awal (Entry age)
Jenis Kelamin (Gender)
Produk (Product)



METHOD

Measurement Scale of Data

	Different	Order	Interval	Absolute Zero
Ratio				
Interval				
Ordinal				
Nominal				

Interval, Ratio = Metrics

Nominal, Ordinal = Non Metrics



METHOD

Appropriate Method base on Measurement Scale

	Dependent	Independent	Interdependent
Classification	NM	M, NM	
Clustering			M, NM
Regression	M	M (+, NM)	
Anova	M	NM (+, M)	
Existing Data	M	NM, M	

Supervised

Unsupervised



METHOD

Linear Regression

Linear regression is used to predict the value of an outcome of Y based on one or more input predictor of variables X.

$$Y = \beta_1 + \beta_2 X + \epsilon$$

Where :

Y = dependent variable

X = independent variable

β_1 = the intercept

β_2 = the slope

ϵ = error term



R PROGRAM SYNTAX

```
> #Inconsistency Data
> usia=df_individu_p$usia.Awal==df_individu_p$usia.Berjalan
> table(usia)
usia
TRUE
7784
```

```
> #Redundancy
> polis=table(df_individu_p$Nomor.Polis)
> polis[polis>1]
named integer(0)
```





R PROGRAM SYNTAX

```
> summary(df_individu_p)
```

```

Bulan.Akuntansi Tahun.Underwriting Bulan.Underwriting Line.Bisnis
Min. : 1.000 Min. :2013 Min. :2006 Min. : 1.00 GROUP : 0
1st Qu.: 4.000 1st Qu.:2014 1st Qu.:2013 1st Qu.: 1.00 INDIVIDUAL:7784
Median : 8.000 Median :2014 Median :2014 Median : 2.00
Mean : 6.818 Mean :2014 Mean :2013 Mean : 3.15
3rd Qu.: 9.000 3rd Qu.:2015 3rd Qu.:2015 3rd Qu.: 4.00
Max. :12.000 Max. :2017 Max. :2017 Max. :12.00

Kelas.Bisnis Currency Nomor.Polis Pekerjaan Tanggal.Masuk
MEDICAL : 98 IDR:7784 Min. : 10 : 3 3/28/2014 0:00 : 33
NON MEDICAL:7686 1st Qu.: 68436 Lain-lain :7769 4/30/2014 0:00 : 27
Median :109747 OWNER : 4 4/30/2013 0:00 : 26
Mean :144028 P.Swasta(Karyawan/i): 0 12/28/2012 0:00: 25
3rd Qu.:174283 Pedagang : 1 2/28/2014 0:00 : 25
Max. :638941 Wiraswasta : 7 4/28/2014 0:00 : 25
(Other) :7623

Tanggal.Lahir Usia.Awal Usia.Berjalan Jenis.Kelamin Smoking.Status
: 373 Min. :20.00 Min. :20.00 : 0 NOT SMOKING:7784
1/1/1975 0:00 : 11 1st Qu.:31.00 1st Qu.:31.00 FEMALE: 55 SMOKING : 0
12/31/1965 0:00: 8 Median :37.00 Median :37.00 MALE :7729
12/31/1967 0:00: 7 Mean :37.48 Mean :37.48
12/31/1975 0:00: 7 3rd Qu.:43.00 3rd Qu.:43.00
1/1/1970 0:00 : 6 Max. :66.00 Max. :66.00
(Other) :7372

Extra.Mortality Produk Lama.Pertanggungan..Tahun. Lama.Pertanggungan..Bulan.
Min. : 0.0000 Produk 38:2638 Min. : 1.000 Min. :0
1st Qu.: 0.0000 Produk 78:2089 1st Qu.: 3.000 1st Qu.:0
Median : 0.0000 Produk 91:1203 Median : 5.000 Median :0
Mean : 0.6115 Produk 82: 251 Mean : 6.745 Mean :0
3rd Qu.: 0.0000 Produk 39: 229 3rd Qu.:10.000 3rd Qu.:0
Max. :200.0000 Produk 68: 209 Max. :28.000 Max. :0
NA's :238 (Other) :1165

Tanggal.Berakhir Uang.Pertanggungan Status
6/28/2016 : 16 Min. :1.200e+06 : 0
9/28/2016 : 16 1st Qu.:6.200e+07 A:7304
10/28/2015: 15 Median :1.350e+08 K: 6
3/28/2016 : 15 Mean :2.326e+08 T: 474
3/28/2017 : 15 3rd Qu.:2.792e+08

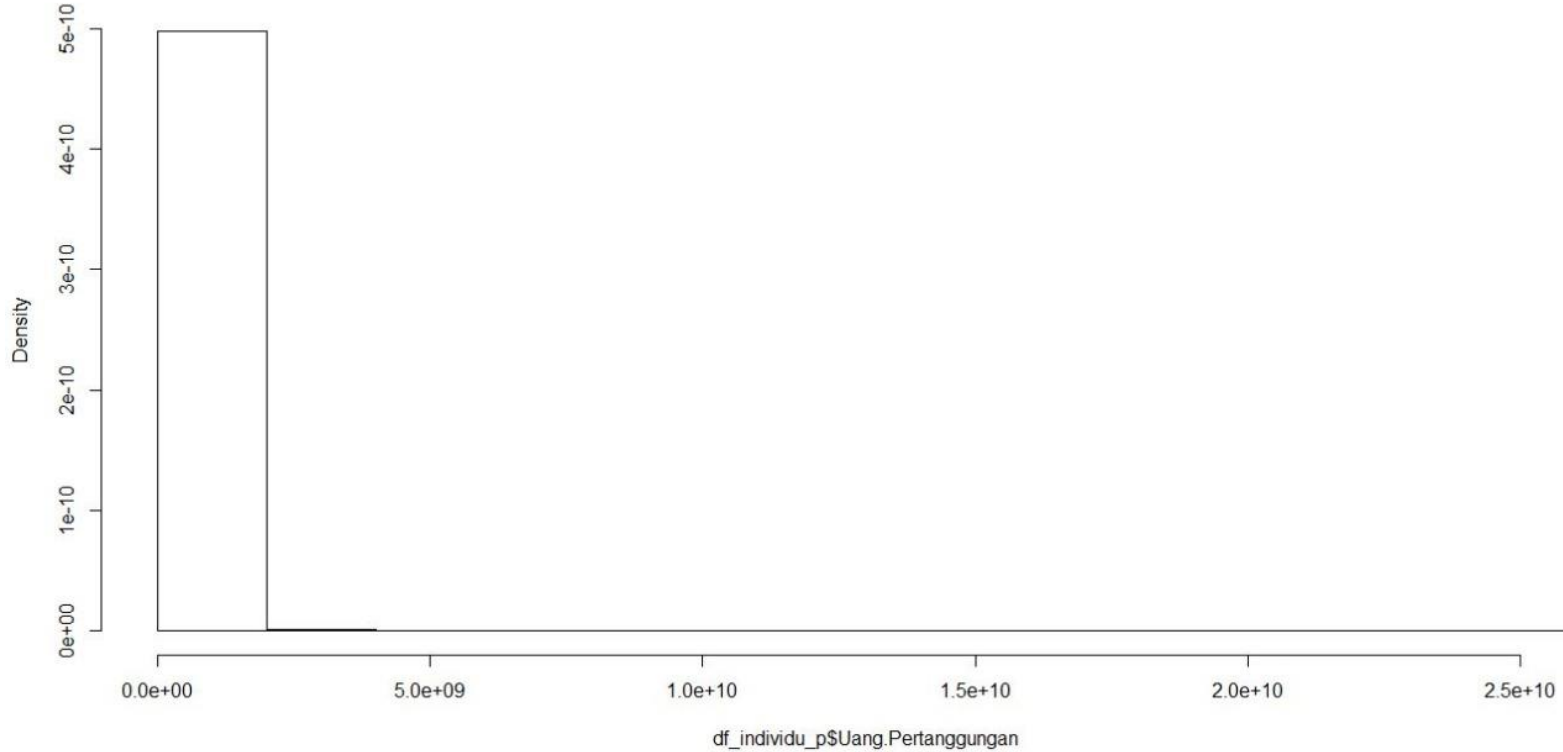
```





R PROGRAM SYNTAX

Histogram of df_individu_p\$Uang.Pertanggung





> summary(linear)

```
call:
lm(formula = Uang.Pertanggunggaan ~ Kelas.Bisnis + Pekerjaan +
  Usia.Awal + Jenis.Kelamin + Produk, data = df_individu3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-802118981 -108088829 -55530783  48372628 1838748997
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 223444948  48343469  4.622 3.86e-06 ***
Kelas.BisnisNON MEDICAL -329547260  27047801 -12.184 < 2e-16 ***
PekerjaanOWNER 405883637  108116897  3.754 0.000175 ***
PekerjaanPedagang 989758710  212868742  4.650 3.38e-06 ***
PekerjaanWiraswasta 959790973  89523098  10.721 < 2e-16 ***
Usia.Awal 3162174  301811  10.477 < 2e-16 ***
Jenis.KelaminMALE -16427872  29571878 -0.556 0.578553
ProdukProduk 15 88975722  69463828  1.281 0.200270
ProdukProduk 16 119529766  106707649  1.120 0.262680
ProdukProduk 19 903949127  46072949  19.620 < 2e-16 ***
ProdukProduk 20 43394552  210212273  0.206 0.836459
ProdukProduk 21 878000053  57914522  15.160 < 2e-16 ***
ProdukProduk 22 1335016286  149460804  8.932 < 2e-16 ***
ProdukProduk 23 119303981  39930441  2.988 0.002819 **
ProdukProduk 24 67757678  95942080  0.706 0.480063
ProdukProduk 25 256325490  96634426  2.653 0.008006 **
ProdukProduk 26 447539931  28860702  15.507 < 2e-16 ***
ProdukProduk 28 601280620  31408806  19.144 < 2e-16 ***
ProdukProduk 30 165812541  66519539  2.493 0.012699 *
ProdukProduk 31 952367592  210251244  4.530 6.00e-06 ***
ProdukProduk 34 358847294  149401399  2.402 0.016334 *
ProdukProduk 36 56016147  122511647  0.457 0.647518
ProdukProduk 37 341209682  95800556  3.562 0.000371 ***
ProdukProduk 38 160456422  21642546  7.414 1.36e-13 ***
ProdukProduk 39 185709044  25446488  7.298 3.21e-13 ***
ProdukProduk 40 87376944  149335357  0.585 0.558494
ProdukProduk 41 172887684  34587040  4.999 5.90e-07 ***
```

First regression model

```
ProdukProduk 42 195632378 210207294 0.931 0.352057
ProdukProduk 44 382516286 210285884 1.819 0.068946 .
ProdukProduk 45 176871500 210363163 0.841 0.400491
ProdukProduk 50 117270382 73022815 1.606 0.108328
ProdukProduk 52 676862436 36003085 18.800 < 2e-16 ***
ProdukProduk 53 767624305 69380519 11.064 < 2e-16 ***
ProdukProduk 54 153556030 39468244 3.891 0.000101 ***
ProdukProduk 55 214653869 52410795 4.096 4.25e-05 ***
ProdukProduk 57 106143385 122774935 0.865 0.387320
ProdukProduk 59 266444594 122546311 2.174 0.029718 *
ProdukProduk 60 53691939 210267699 0.255 0.798459
ProdukProduk 65 54340514 81738335 0.665 0.506192
ProdukProduk 67 184467418 40265145 4.581 4.69e-06 ***
ProdukProduk 68 134537644 25685144 5.238 1.67e-07 ***
ProdukProduk 70 126363350 106825576 1.183 0.236888
ProdukProduk 73 246179805 122572019 2.008 0.044630 *
ProdukProduk 75 284205419 210243666 1.352 0.176482
ProdukProduk 76 217101893 46227003 4.696 2.69e-06 ***
ProdukProduk 78 202052829 21513056 9.392 < 2e-16 ***
ProdukProduk 80 87763410 149423335 0.587 0.556988
ProdukProduk 81 151084475 29338652 5.150 2.67e-07 ***
ProdukProduk 82 138043436 25118984 5.496 4.02e-08 ***
ProdukProduk 83 174544572 55118641 3.167 0.001548 **
ProdukProduk 84 99205419 210243666 0.472 0.637041
ProdukProduk 87 241280951 27336081 8.826 < 2e-16 ***
ProdukProduk 88 857167592 210251244 4.077 4.61e-05 ***
ProdukProduk 90 269093013 26181497 10.278 < 2e-16 ***
ProdukProduk 91 313471577 21791286 14.385 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 209200000 on 7699 degrees of freedom
Multiple R-squared: 0.266, Adjusted R-squared: 0.2609
F-statistic: 51.67 on 54 and 7699 DF, p-value: < 2.2e-16
```


Final Regression Model

```
> summary(linear6)
```

```
Call:
```

```
lm(formula = Uang.Pertanggungaban ~ Kelas.Bisnis + Pekerjaan +  
  Usia.Awal + Produk, data = df_individu3)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-801836193	-108375197	-57220728	49309409	1839414851

```
Coefficients:
```

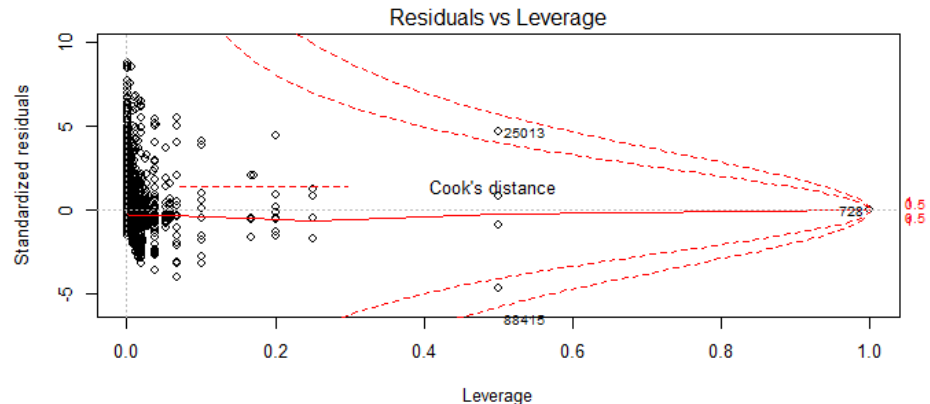
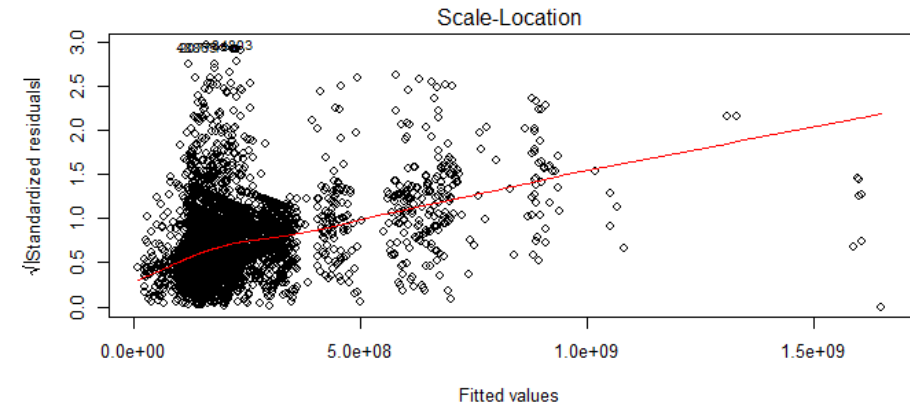
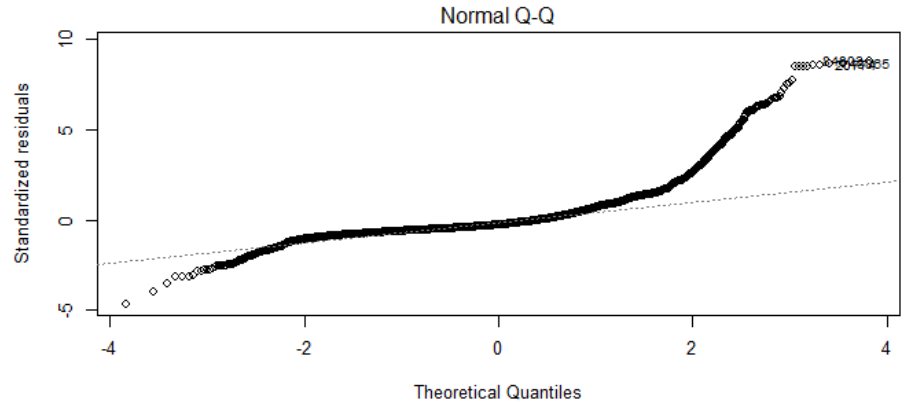
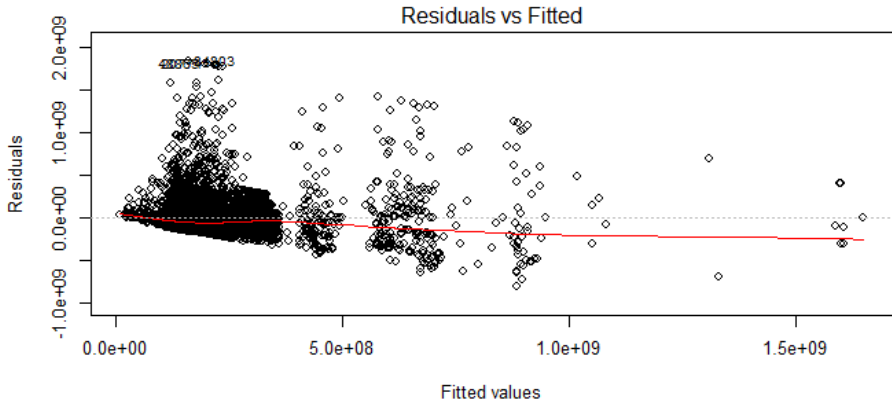
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	283262633	34278152	8.264	< 2e-16 ***
Kelas.BisnisNON MEDICAL	-331371649	27048447	-12.251	< 2e-16 ***
PekerjaanOWNER	408583606	107920220	3.786	0.000154 ***
PekerjaanPedagang	1004093408	210920782	4.761	1.97e-06 ***
PekerjaanWiraswasta	960408313	89432817	10.739	< 2e-16 ***
Usia.Awal	2912655	295610	9.853	< 2e-16 ***
ProdukProduk 19	839230387	43483602	19.300	< 2e-16 ***
ProdukProduk 21	813526352	55897517	14.554	< 2e-16 ***
ProdukProduk 22	1269078744	148685422	8.535	< 2e-16 ***
ProdukProduk 25	197607844	94694776	2.087	0.036940 *
ProdukProduk 26	382611602	24473127	15.634	< 2e-16 ***
ProdukProduk 28	536054667	27394046	19.568	< 2e-16 ***
ProdukProduk 31	887428124	209742619	4.231	2.35e-05 ***
ProdukProduk 34	294157344	148656054	1.979	0.047876 *
ProdukProduk 37	277867132	94651628	2.936	0.003338 **
ProdukProduk 38	95100618	15140553	6.281	3.54e-10 ***
ProdukProduk 39	119947665	20157368	5.951	2.79e-09 ***
ProdukProduk 41	107995538	31027227	3.481	0.000503 ***
ProdukProduk 52	612620282	32631364	18.774	< 2e-16 ***
ProdukProduk 53	703084066	67707537	10.384	< 2e-16 ***
ProdukProduk 54	87213021	36246161	2.406	0.016146 *
ProdukProduk 55	149622473	50130296	2.985	0.002848 **
ProdukProduk 67	119831418	37279282	3.214	0.001312 **
ProdukProduk 68	69026313	20528655	3.362	0.000776 ***
ProdukProduk 76	151365885	43562568	3.475	0.000514 ***
ProdukProduk 78	137150884	15126814	9.067	< 2e-16 ***

ProdukProduk 78	137150884	15126814	9.067	< 2e-16	***
ProdukProduk 81	85478069	24937111	3.428	0.000612	***
ProdukProduk 82	72124782	19740340	3.654	0.000260	***
ProdukProduk 83	108196058	52865532	2.047	0.040728	*
ProdukProduk 87	176598030	22693081	7.782	8.06e-15	***
ProdukProduk 88	792228124	209742619	3.777	0.000160	***
ProdukProduk 90	205335668	21429501	9.582	< 2e-16	***
ProdukProduk 91	249050413	15619700	15.945	< 2e-16	***

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

```
Residual standard error: 209200000 on 7721 degrees of freedom  
Multiple R-squared:  0.2636,    Adjusted R-squared:  0.2605  
F-statistic: 86.36 on 32 and 7721 DF,  p-value: < 2.2e-16
```

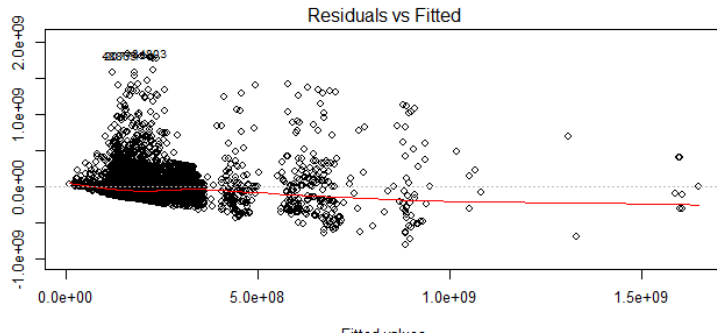
Plot of Model without Outlier





SUMMARY

1. Influencing variables are :
 - Kelas Bisnis (business class)
 - Pekerjaan (Job)
 - Usia Awal (Entry age)
 - Produk (Product)
2. Goodness of fit





Thank you

Plot of Model with Outlier

