

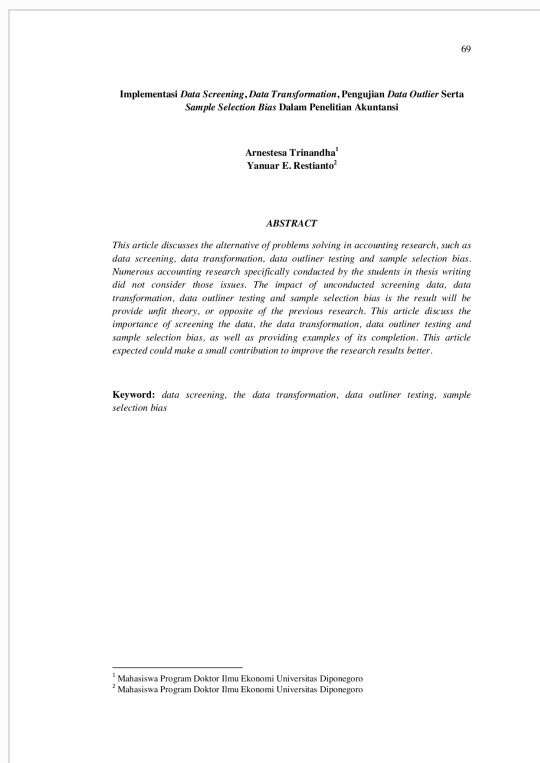


Digital Receipt

This receipt acknowledges that **Turnitin** received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Yanuar Eko Restianto
Assignment title: DOSEN
Submission title: Implementasi Data Screening, Data...
File name: 2_Yanuar_-_Akuntabel_Mei_2014.pdf
File size: 1.44M
Page count: 13
Word count: 2,035
Character count: 12,453
Submission date: 26-Nov-2019 01:42PM (UTC+0700)
Submission ID: 1221995314



Implementasi Data Screening, Data Transformation, Pengujian Data Outlier Serta Sample Selection Bias Dalam Penelitian Akuntansi

by Yanuar Eko Restianto

Submission date: 26-Nov-2019 01:42PM (UTC+0700)

Submission ID: 1221995314

File name: 2_Yanuar_-_Akuntabel_Mei_2014.pdf (1.44M)

Word count: 2035

Character count: 12453

**Implementasi *Data Screening*, *Data Transformation*, *Pengujian Data Outlier* Serta
Sample Selection Bias Dalam Penelitian Akuntansi**

**Arnestesa Trinandha¹
Yanuar E. Restianto²**

ABSTRACT

This article discusses the alternative of problems solving in accounting research, such as data screening, data transformation, data outlier testing and sample selection bias. Numerous accounting research specifically conducted by the students in thesis writing did not consider those issues. The impact of uncondacted screening data, data transformation, data outlier testing and sample selection bias is the result will be provide unfit theory, or opposite of the previous research. This article discuss the importance of screening the data, the data transformation, data outlier testing and sample selection bias, as well as providing examples of its completion. This article expected could make a small contribution to improve the research results better.

Keyword: *data screening, the data transformation, data outlier testing, sample selection bias*

¹ Mahasiswa Program Doktor Ilmu Ekonomi Universitas Diponegoro

² Mahasiswa Program Doktor Ilmu Ekonomi Universitas Diponegoro

I. Pendahuluan

Banyak penelitian akuntansi khususnya yang dilakukan oleh para mahasiswa dalam menyusun skripsi atau tesis mengesampingkan masalah data *screening*, *data transformation*, pengujian *data outlier* serta *sample selection bias*. Penting untuk mempertimbangkan kualitas data, karena karakteristik data akan berpengaruh terhadap *output* yang dihasilkan (Meyers, 2005). Dampak dari tidak dilakukannya hal tersebut adalah hasil penelitian yang biasanya tidak sesuai dengan teori, atau berlawanan dengan hasil-hasil penelitian sebelumnya. Setidaknya terdapat empat prosedur untuk mengatasi hal tersebut yakni: 1) data *screening* khususnya untuk memastikan data yang akan diuji adalah data yang normal; 2) *data transformation*, digunakan untuk mentransformasi data yang tidak normal menjadi normal; 3) pengujian *data outlier* dan 4) *sample selection bias*, digunakan untuk memastikan bahwa model penelitian bebas dari bias pemilihan sampel.

II. Data Screening

Sebelum melakukan uji statistik, terlebih dahulu dilakukan *screening* terhadap data yang akan diolah, fungsinya adalah memastikan bahwa data telah memenuhi asumsi normalitas. Wulder (1998) menjelaskan bahwa keputusan untuk melakukan data *screening* akan mempengaruhi keputusan untuk langkah-langkah statistik berikutnya. Misal, jika data tidak normal dan memiliki *outlier*, akan dihadapkan pada keputusan berikutnya untuk menghapus atau mengubah data. Jika keputusan pertama berikutnya adalah melakukan transformasi data, ada kemungkinan data *outlier* akan menjadi

lebih sedikit, namun jika data *outlier* dihapus atau diubah terlebih dahulu, ada kemungkinan akan lebih sedikit data yang tidak normal. Transformasi data biasanya lebih disukai karena biasanya mengurangi jumlah *outlier* dan lebih memungkinkan untuk menghasilkan normalitas, linearitas dan homokedastisitas. Jika asumsi ini dipenuhi, maka nilai residual dari analisis juga berdistribusi normal dan independen.

- *Screening* data dapat dilakukan dengan uji normalitas dengan melihat nilai *skewness* dan kurtosis, melihat nilai signifikansi pada uji *kolmogorov smirnov*, dan melihat untuk grafik histogram.
- *Skewness* berhubungan dengan simetri distribusi. *Skewed variable* (variabel menceng) adalah variabel yang nilai *mean*-nya tidak di tengah-tengah distribusi. Sedangkan kurtosis berhubungan dengan puncak dari suatu distribusi. Jika variabel terdistribusi secara normal maka nilai *skewness* dan kurtosis sama dengan nol. Terdapat uji signifikansi *skewness* dan kurtosis dengan cara sebagai berikut (Ghozali, 2013):

$$Z_{skew} = \frac{S - 0}{\frac{\sqrt{6}}{N}}$$

$$Z_{kurt} = \frac{K - 0}{\sqrt{24/N}}$$

dimana:

- S : nilai *skewness*
- K : nilai kurtosis
- N : jumlah kasus

Langkah dan Contoh Analisis

1. Uji *Skewness* dan Kurtosis

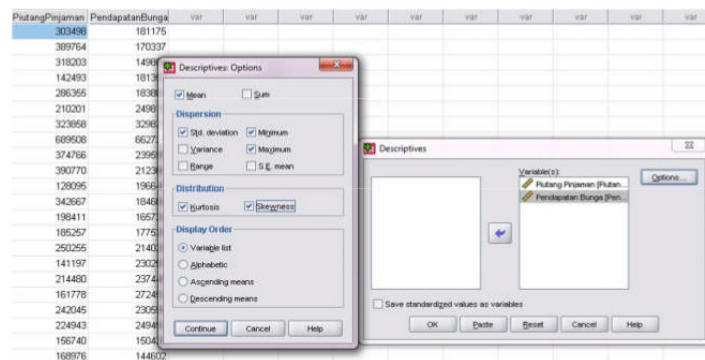
- Terdapat 100 sampel data piutang pinjaman dan

pendapatan bunga, akan diuji normalitas untuk melakukan *screening*

terhadap data yang akan diuji.

	PiutangPinjaman	PendapatanBunga
1	303498	181175
2	389764	170337
3	318203	149864
4	142493	181352
5	286355	183803
6	210201	249810
7	323858	329874
8	689508	662724
9	374766	239598
10	390770	212366
11	128095	196645
12	342667	184600
13	198411	165726
14	185257	177538
15	250255	214025
16	141197	230292

- Uji nilai *skewness* dan *kurtosis* pada *Analysis-Descriptive Statistics* (Ghozali, 2013)



- Output yang dihasilkan sebagai berikut :

2 Descriptive Statistics

	N	Min	Max	Mean	Std. Deviation	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std. Error	Stat	Std. Error
Pendapatan Bunga	100	49128	2185835	343816.57	317398.877	3.659	.241	16.350	.478
Piutang Pinjaman	100	69312	1186603	316865.85	189468.646	2.035	.241	6.062	.478
Valid N (listwise)	100								

Berdasarkan hasil tersebut, dihitung nilai Z_{skew} dan Z_{kurt} :

$$Z_{skew}(PB) = \frac{3.659}{\frac{\sqrt{6}}{100}} = 14.93$$

$$Z_{kurt}(PB) = \frac{16.350}{\frac{\sqrt{24}}{100}} = 33.37$$

$$Z_{skew}(PP) = \frac{2.035}{\frac{\sqrt{6}}{100}} = 8.30$$

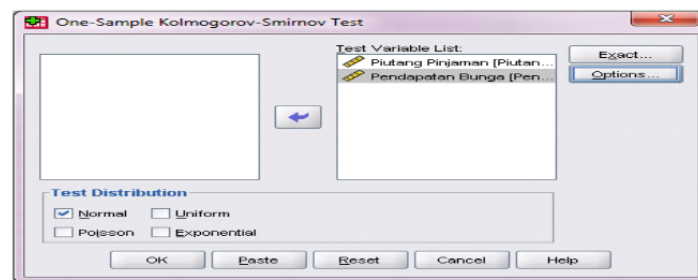
$$Z_{kurt}(PP) = \frac{6.062}{\frac{\sqrt{24}}{100}} = 12.37$$

Hasil perhitungan *Z skewness* dan kurtosis untuk variabel pendapatan bunga (PB) dan piutang pinjaman (PP) menghasilkan nilai yang jauh di atas nilai kritisnya yaitu ± 2.58 (signifikan pada $\alpha=0.01$). Jadi dapat disimpulkan bahwa kedua

variabel PP dan PB tidak terdistribusi secara normal.

2. Uji Kolmogorov Smirnov

- Uji *kolmogorov smirnov* dilakukan untuk menguji normalitas data dengan merumuskan hipotesis terlebih dahulu:
 H_0 = data terdistribusi secara normal
 H_a = data tidak terdistribusi secara normal
- Uji dengan SPSS *Analyze-Non Parametric Tests-1 Sample K-S* (Ghozali, 2013):



- Hasil pengujian sebagai berikut :

NPar Tests

[DataSet2]

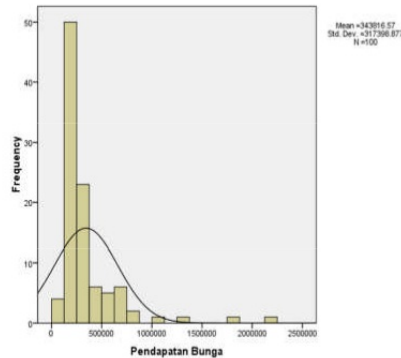
One-Sample Kolmogorov-Smirnov Test				
		Piutang Pinjaman	Pendapatan Bunga	
N		100	100	
Normal Parameters ^{a,b}	Mean	316865.85	343816.57	
	Std. Deviation	189468.646	317398.877	
Most Extreme Differences	Absolute	.137	.248	
	Positive	.137	.248	
	Negative	-.111	-.215	
Kolmogorov-Smirnov Z		1.373	2.479	
Asymp. Sig. (2-tailed)		.046	.000	
a. Test distribution is Normal.				
b. Calculated from data.				

- Nilai K-S pada variabel Piutang Pinjaman sebesar 1.373 dengan probabilitas 0.046 (di bawah $\alpha=0.05$),

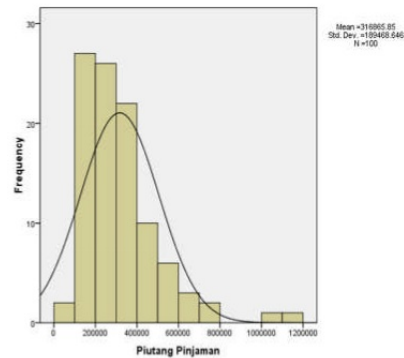
dan nilai K-S pada pendapatan bunga sebesar 2.479 dengan probabilitas 0.000 (di bawah $\alpha=0.05$),

sehingga ¹ hipotesis nol ditolak atau variabel piutang pinjaman dan pendapatan bunga tidak terdistribusi secara normal.

3. Uji Normalitas dengan Grafik Histogram



- Uji normalitas dapat dilakukan dengan menggunakan grafik histogram, grafik ini juga dapat digunakan untuk menentukan pola transformasi data (Ghozali, 2013).

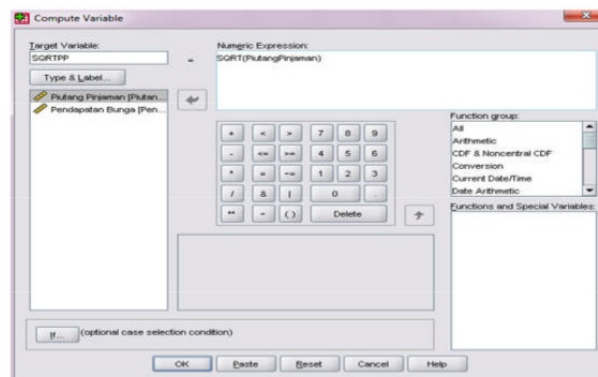


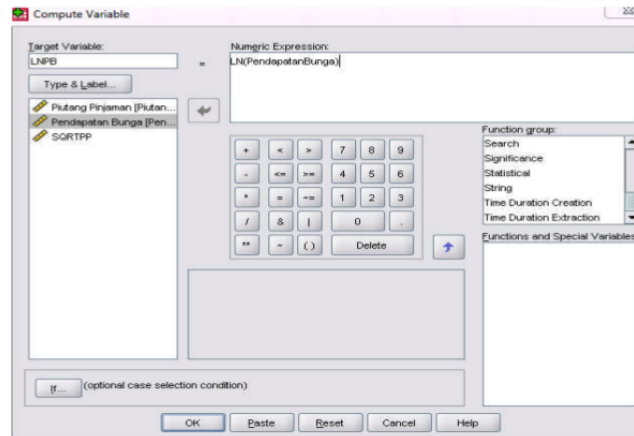
III. Data Transformation

- Prosedur transformasi data dapat digunakan untuk memodifikasi variabel yang melanggar asumsi statistik normalitas, linieritas dan homokedastisitas, atau yang memiliki pola *outlier* yang tidak biasa (Hair et.al., 1998).
- Berdasarkan hasil uji normalitas dengan grafik histogram, menunjukkan

bahwa ketidaknormalan data termasuk dalam jenis *Moderate Positive Skewness*, sehingga data tersebut dapat ditransformasikan dengan menggunakan $\sqrt{x}$ atau akar kuadrat atau logaritma 10 atau LN.

- Pengujian dengan SPSS :
Transform – Compute Variable





- Setelah ditransformasi data, maka *data view* akan muncul 2 variabel baru yakni SQRTTP dan LNPB

PiutangPinjaman	PendapatanBunga	SQRTTP	LNPB
303498	181175	550.91	12.11
389764	170337	624.31	12.05
318203	149864	564.09	11.92
142493	181352	377.48	12.11
266355	183803	535.12	12.12
210201	249810	458.48	12.43
323858	329874	569.09	12.71
689508	662724	830.37	13.40
374766	239598	612.18	12.39
390770	212366	625.12	12.27
128095	196645	357.90	12.19
342667	184600	585.38	12.13
198411	165726	445.43	12.02
185257	177538	430.41	12.09
250255	214025	500.25	12.27
141197	230292	375.76	12.35

- Selanjutnya dilakukan *smirnov* dan grafik kembali uji normalitas data atas variabel SQRTTP dan LNPB dengan menggunakan *skewness* dan *kurtosis*; uji *kolmogorov* dan grafik histogram.
- Hasil pengujian setelah ditransformasi dengan SQRTTP dan LNPB:

2 Descriptive Statistics

	N	Min	Max	Mean	Std. Dev	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std. Error	Stat	Std. Error
SQRTTP	100	263.27	1089.31	542.5906	150.62584	1.011	.241	1.736	.478
LNPB	100	10.80	14.60	12.5241	.61089	.822	.241	1.933	.478
Valid N (listwise)	100								

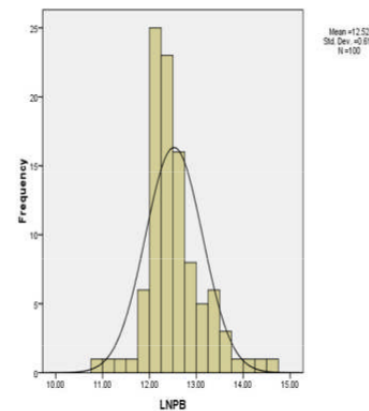
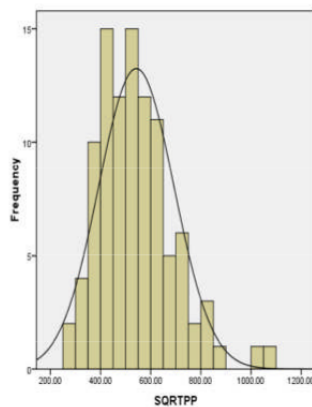
4

One-Sample Kolmogorov-Smirnov Test

		SQRTTP	LNPB
N		100	100
Normal Parameters ^{a,b}	Mean	542.5906	12.5241
	Std. Deviation	150.62584	.61089
Most Extreme Differences	Absolute	.081	.132
	Positive	.081	.132
	Negative	-.064	-.104
Kolmogorov-Smirnov Z		.808	1.316
Asymp. Sig. (2-tailed)		.531	.063

a. Test distribution is Normal.

b. Calculated from data.



- Hasil uji *kolmogorov smirnov* menunjukkan bahwa nilai K-S variabel SQRTTP sebesar 0.808 dengan probabilitas 0.531 ($> \alpha=0.05$) dan nilai K-S variabel LN-PB sebesar 1.316 dengan probabilitas sebesar 0.063 ($> \alpha=0.05$), sehingga dapat disimpulkan data kedua variabel tersebut telah terdistribusi dengan normal.

IV. Data Outlier

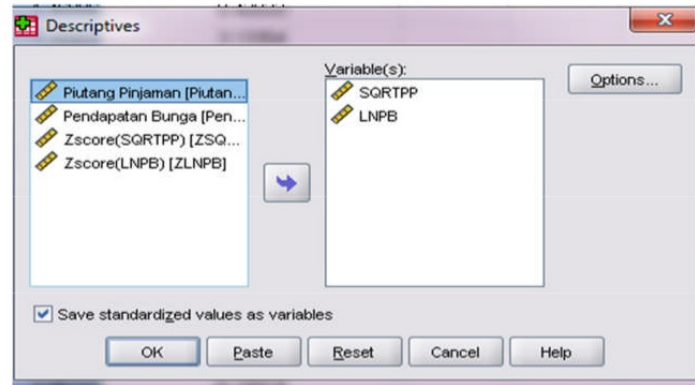
- Data outlier dapat didefinisikan sebagai data yang muncul memiliki karakteristik unik yang terlihat sangat jauh berbeda dari observasi-observasi lainnya dan

muncul dalam bentuk nilai yang ekstrim, baik untuk sebuah variabel tunggal atau variabel kombinasi. (Hair et.al., 1998).

- Menurut Hair et.al. (1998) untuk kasus sampel kecil (kurang dari 80), maka standar skor dengan nilai ≥ 2.5 dinyatakan *outlier*. Untuk sampel ukuran besar dinyatakan *outlier* jika nilainya berkisar antara 3 sampai 4. Jika standar skor tidak digunakan, maka dapat ditentukan data *outlier* yang nilainya lebih besar dari 2.5 (untuk sampel kecil) standar deviasi, atau antara 3 sampai 4 (untuk sampel besar) standar deviasi.
- Data *outlier* dapat dilihat dengan membuat variabel baru

ZSQRTPP dan ZLNBP melalui
SPSS *Analyze-Descriptive*

Statistics-Descriptives :



- Dari hasil uji deskriptif akan muncul 2 variabel ZSQRTPP dan ZLNBP. Jika terdapat nilai Z di kisaran 3-4 (data di atas 80), maka data tersebut adalah *outlier*.
- Hasil di bawah ini menunjukkan ada 2 data outlier pada variabel ZSQRT dan 2 data *outlier* pada variabel SLNPB.

	Piutang Pinjaman [Piutan...	Pendapatan Bunga [Pen...	SQRTPP	LNBP	Z SQRTPP	ZLNBP
1	1186603	2185835	1089.31	14.60	3.62967	3.39417
2	1093675	1044764	1045.79	13.86	3.34072	2.18575
3	786062	532927	886.60	13.19	2.20338	1.08381
4	709457	329243	842.29	12.70	1.98971	0.29546
5	689508	662724	830.37	13.40	1.91053	1.44062
6	670510	761834	818.85	13.54	1.83406	1.66877
7	620431	316762	787.67	12.67	1.62710	0.23220
8	583497	355341	763.87	12.78	1.46906	0.42033
9	556427	1864111	745.94	14.44	1.35003	3.13354
10	536260	721227	732.30	13.49	1.25946	1.57910
11	533404	1363127	730.35	14.13	1.24650	2.62116
12	512104	507737	715.61	13.14	1.14870	1.00454
13	500128	587204	707.20	13.28	1.09282	1.24257
14	492487	205434	701.77	12.24	1.05681	-0.46870
15	463659	414743	680.93	12.94	0.91840	0.67338
16	452246	467636	679.89	13.06	0.91150	0.86986
17	438407	660883	662.12	13.40	0.79357	1.43607
18	437971	219436	661.79	12.30	0.79139	-0.36871
19	432201	630676	657.42	13.35	0.76235	1.35948
20	422439	792751	649.95	13.58	0.71278	1.73388
21	413529	258465	643.06	12.46	0.66703	-0.10074
22	412120	378641	641.97	12.84	0.65975	0.52430
23	401721	247107	633.81	12.42	0.60563	-0.17431
24	399334	663172	631.93	13.40	0.59311	1.44173
25	397223	228916	630.26	12.34	0.58201	-0.29948

	PiutangPinjaman	PendapatanBunga	SortTPP	LNBP	ZSortTPP	ZLNBP
1	1186603	2185835	1089.31	14.60	3.62967	3.39417
2	556427	1864111	745.94	14.44	1.35003	3.13354
3	533404	1363127	730.35	14.13	1.24650	2.62116
4	1093675	1044764	1045.79	13.86	3.34072	2.18575
5	422439	792751	649.95	13.58	0.71278	1.73388
6	670510	761834	818.85	13.54	1.83406	1.66877
7	324925	739544	570.02	13.51	0.18212	1.62016
8	536260	721227	732.30	13.49	1.25946	1.57910
9	399334	663172	631.93	13.40	0.69311	1.44173
10	689508	662724	830.37	13.40	1.91053	1.44062
11	438407	660883	662.12	13.40	0.79357	1.43607
12	432201	630676	657.42	13.35	0.76235	1.35948
13	500128	587204	707.20	13.28	1.09282	1.24257
14	786062	532927	886.60	13.19	2.28388	1.08381
15	202041	525091	449.49	13.17	-0.61809	1.05955
16	302304	524341	549.82	13.17	0.04801	1.05722
17	512104	507737	715.61	13.14	1.14870	1.00454
18	462246	467636	679.89	13.06	0.91150	0.86986
19	265899	422923	515.65	12.95	-0.17883	0.70535
20	463659	414743	680.93	12.94	0.91840	0.67338
21	200196	390318	447.43	12.87	-0.63175	0.57402
22	229447	379929	479.01	12.85	-0.42213	0.52986
23	412120	378641	641.97	12.84	0.65975	0.52430
24	213088	365172	461.61	12.81	-0.53760	0.46501
25	277071	358721	526.38	12.79	-0.10765	0.43583

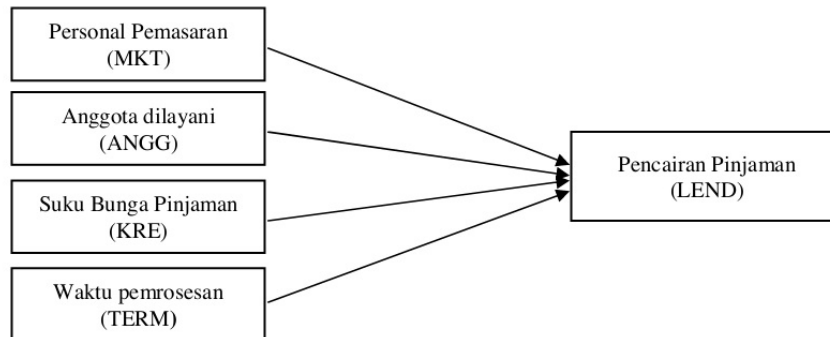
V. Sample Selection Bias

Heckman (1979) menjelaskan bahwa *sample selection bias* berkaitan dengan bias yang disebabkan pemilihan sampel non random untuk mengestimasi hubungan perilaku sebagai kesalahan spesifikasi yang biasa timbul karena masalah data yang hilang. Analisis pemilihan sampel bias digunakan untuk memperkirakan variabel yang ketika dihilangkan dari analisis regresi akan menimbulkan kesalahan spesifikasi. Pemilihan sampel bias timbul dalam praktik karena dua alasan : pertama, karena adanya pemilihan tersendiri oleh individu atau unit data yang sedang diinvestigasi; kedua, karena keputusan pemilihan sampel oleh analis atau pengolah data dioperasikan dengan banyak cara, yang sama seperti pemilihan sendiri.

Ilustrasi Sample Selection Bias :

Penelitian ini bertujuan untuk menguji pengaruh personal pemasaran (MKT), jumlah anggota dilayani (ANGG), Suku bunga pinjaman (KRE) dan waktu pemrosesan pinjaman (TERM) terhadap jumlah pencairan pinjaman (LEND). Sebelum dilakukan uji statistik, maka dilakukan *sample selection bias* dengan tujuan untuk melihat apakah variabel dependen tersedia untuk semua responden dan distribusi responden atas pengkategorian variabel independen dilakukan secara benar atau tidak bias. Seleksi bias merupakan ancaman bagi validitas internal dalam variabel independen yang berkorelasi dengan error (Cuddeback, 2004).

Kerangka Penelitian :



Hipotesis:

Berdasarkan kerangka penelitian, dapat dirumuskan hipotesis sebagai berikut :

“Personal pemasaran (MKT), jumlah anggota dilayani (ANGG), Suku bunga pinjaman (KRE) dan waktu pemrosesan pinjaman (TERM) berpengaruh terhadap jumlah pencairan pinjaman (LEND)”.

$$\text{LEND} = a + \beta_1 \text{MKT} + \beta_2 \text{NPL} + \beta_3 \text{KRE} + \beta_4 \text{TERM} + \varepsilon$$

Keterangan:

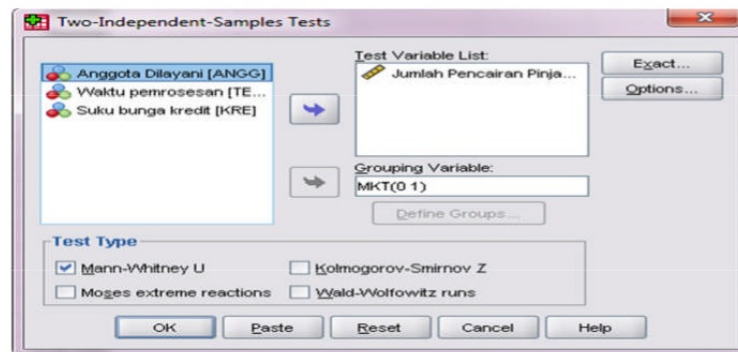
MKT 0 = Laki - laki
1 = Perempuan

- 2) Melakukan uji beda terhadap variabel independen dummy, yaitu variabel MKT. Uji beda dilakukan dengan *Mann Whitney U* pada SPSS (Ghozali, 2013).

Langkah Analisis :

- 1) Membuat persamaan awal dari model penelitian

Persamaan 1 :



3 Output uji *Mann-Whitney U*

Ranks

	Marketing	N	Mean Rank	Sum of Ranks
Jumlah Pencairan Pinjaman	Laki-laki	20	11.63	232.50
	Perempuan	20	29.38	587.50
	Total	40		

Test Statistics^b

	Jumlah Pencairan Pinjaman
3 Mann-Whitney U	22.500
Wilcoxon W	232.500
Z	-4.803
Asymp. Sig. (2-tailed)	.000
Exact Sig. [2*(1-tailed Sig.)]	.000 ^a

a. Not corrected for ties.

b. Grouping Variable: Marketing

Berdasarkan tabel output uji *Mann-Whitney U* menunjukkan bahwa nilai *Asymp.Sig. (2-*

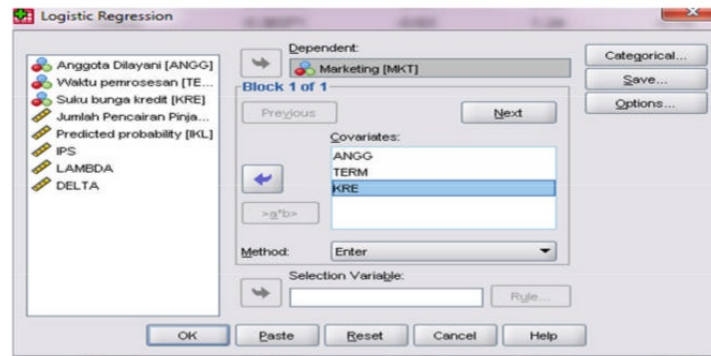
tailed) sebesar 0.000 atau di bawah 0.05. Hal ini menunjukkan bahwa terdapat

perbedaan jumlah pencairan pinjaman (LEND) antara personal pemasaran laki-laki dan perempuan. Perbedaan variabel independen tersebut dapat menyebabkan timbulnya *sample selection bias*, sehingga

langkah berikutnya harus dilakukan transformasi data.

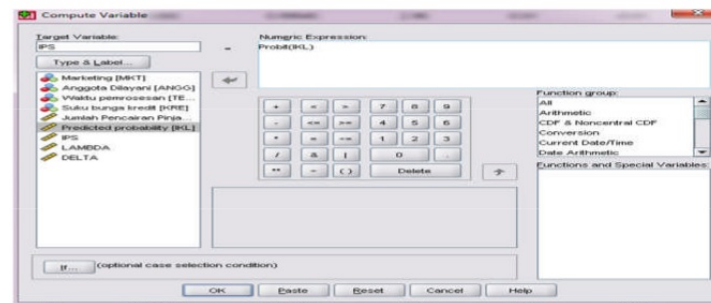
3) Membuat variabel baru IKL

- Melalui SPSS : *Analyze – Regression – Binary Logistic* (Ghozali, 2013)



4) Membuat variabel baru IPS

Membuat variabel *Individual Probit Scores* (IPS) melalui Transform dengan rumus sebagai berikut:
 $IPS = Probit (IKL)$



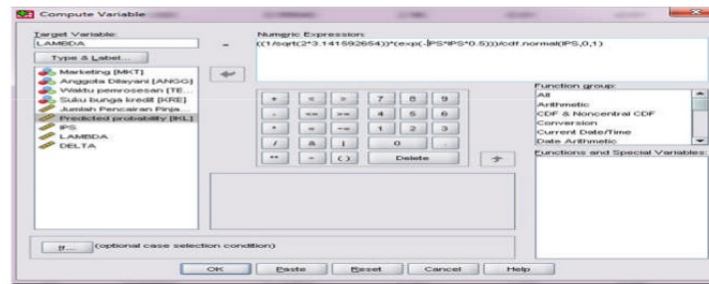
5) Menghitung Lambda

Individual Probit Scores (IPS) dihitung dan ditambahkan pada file data yang diberi nama variabel

“IPS”. Skor *probit* tersebut digunakan untuk menghitung faktor pengendali yang disebut LAMBDA. Heckman (1979)

5

$$LAMBDA = ((1/\sqrt{2*3.141592654}))*(\exp(-IPS*IPS*0.5))/cdf.normal(IPS,0,1)$$



6) Menghitung Delta

Tahap terakhir adalah menghitung DELTA dengan rumus berikut: Heckman (1979)

$$DELTA = -LAMBDA * IPS - LAMBDA * LAMBDA$$

Nilai DELTA harus diantara -1 and 0. Hal ini menunjukkan bahwa LAMBDA telah dihitung dengan baik, dan menunjukkan bahwa data tidak ada bias.

TS-DATA SELEKSI BIAS (Data Set) - SPSS Statistics Data Editor

	MKT	ANGG	TERM	KRE	LEND	IKL	IPS	LAMBDA	DELTA
1	Perempuan	7	3	11	247000	0.97047	1.89	0.07	-0.14
2	Laki-laki	2	2	12	502000	0.56880	0.17	0.69	-0.60
3	Perempuan	11	3	11	3105000	0.99845	2.96	0.01	-0.01
4	Perempuan	12	2	12	4100000	0.99956	3.32	0.00	-0.01
5	Perempuan	10	4	10	3710000	0.99462	2.55	0.02	-0.04
6	Perempuan	9	3	11	2651000	0.99318	2.47	0.02	-0.05
7	Laki-laki	3	4	11	785000	0.26371	-0.63	1.24	-0.75
8	Laki-laki	1	5	12	255000	0.00610	-2.51	2.83	-0.91
9	Laki-laki	2	7	10	378000	0.00467	-2.60	2.91	-0.92
10	Perempuan	11	2	12	3985000	0.99907	3.11	0.00	-0.01
11	Laki-laki	1	4	11	540000	0.07477	-1.44	1.89	-0.85
12	Perempuan	5	3	12	1755000	0.72479	0.60	0.46	-0.49
13	Laki-laki	4	5	12	1057000	0.05419	-1.61	2.03	-0.86
14	Laki-laki	5	6	11	1005000	0.06776	-1.49	1.93	-0.85
15	Perempuan	9	5	13	4150000	0.45683	-0.11	0.87	-0.66
16	Perempuan	8	4	12	3788000	0.84020	1.00	0.29	-0.37
17	Laki-laki	3	6	10	978000	0.04415	-1.70	2.11	-0.87
18	Laki-laki	1	4	11	299000	0.07477	-1.44	1.89	-0.85
19	Laki-laki	2	5	14	311000	0.00163	-2.94	3.23	-0.93
20	Laki-laki	2	5	13	475000	0.00457	-2.61	2.92	-0.92
21	Perempuan	12	2	11	3799000	0.99984	3.60	0.00	0.00
22	Perempuan	10	3	13	2187000	0.97479	1.96	0.06	-0.12
23	Perempuan	15	4	13	3155000	0.99709	2.76	0.01	-0.02
24	Perempuan	2	3	12	489000	0.22015	-0.77	1.35	-0.77
25	Perempuan	3	2	12	950000	0.73524	0.63	0.45	-0.48
26	Laki-laki	4	3	12	1198000	0.55576	0.14	0.71	-0.60
27	Perempuan	5	5	12	2100000	0.10763	-1.24	1.72	-0.83
28	Perempuan	6	3	11	3500000	0.93981	1.55	0.13	-0.21
29	Laki-laki	2	2	13	175000	0.31898	-0.47	1.12	-0.73
30	Laki-laki	1	5	12	240000	0.00610	-2.51	2.83	-0.91
31	Laki-laki	1	5	13	395000	0.00610	-2.51	2.83	-0.91

VI. Kesimpulan

1. Dalam proses pengumpulan data, seringkali peneliti menemukan nilai pengamatan yang bervariasi. Keberagaman data dengan berbagai karakteristik, perlu dilakukan pengujian awal terhadap kualitas data sebelum dilakukan analisa statistik lebih lanjut.
2. Data *screening*, *data transformation*, pengujian *data outlier* serta *sample selection bias* dapat dilakukan untuk menjawab pertanyaan apakah data akurat mencerminkan jawaban oleh peserta; apakah ada pola data yang hilang; apakah ada data yang ekstrim yang mungkin mendistorsi pemahaman peneliti terhadap fenomena yang diteliti; apakah data telah memenuhi asumsi statistik dan apa yang dapat dilakukan jika beberapa data ternyata melanggar asumsi statistik?
3. Dengan demikian, *data screening*, *data transformation*, pengujian *data outlier* serta *sample selection bias* dapat dikatakan sebagai solusi tahap awal untuk menghasilkan data yang layak untuk analisa data uji statistik berikutnya.

DAFTAR PUSTAKA

- 8 Cuddeback Gary et.al. 2004. Detecting and Statistically Correcting Sample Selection Bias. *Journal of Social Service Research*, Vol. 30(3), pp.19-33.
- 7 Gerdin J., Greve J., 2002. Forms of contingency fit in management accounting research-a critical review, *Accounting, Organizations and Society*, 29, pp.303-326.
- Ghozali, Imam. 2013. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 21*. Badan Penerbit Universitas Diponegoro.
- 6 Hahn, E.D., Soyer, Refik. 2005. Probit and Logit Models: Differences in the Multivariate. *The Journal of the Royal Statistical Society, Series B*.
- 9 James J. Heckman. 1979. Sample Selection Bias as a Specification Error, *Econometrica*, Vol. 47, No. 1 pp. 153-161.
- Leone, Andrew J., Minutti-Meza, Miguel dan Wasley, Charles E. 2014. Influential Observations and Inference in Accounting Research. *Simon Business School Working Paper No. FR 14-06*.
- Meyers S. Lawrence et.al. 2005. *Applied Multivariate Research: Design and Interpretation*. Sage Publications, Inc.
- Wulder Mike. 1998. A Practical Guide to the Use of Selected Multivariate Statistics. *Pacific Forestry Centre*. Canada.

Implementasi Data Screening, Data Transformation, Pengujian Data Outlier Serta Sample Selection Bias Dalam Penelitian Akuntansi

ORIGINALITY REPORT

9%

SIMILARITY INDEX

6%

INTERNET SOURCES

3%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universitas Pelita Harapan

Student Paper

2%

2

hdl.handle.net

Internet Source

2%

3

www.zn-k.ru

Internet Source

2%

4

lrd.yahooapis.com

Internet Source

1%

5

home.planet.nl

Internet Source

1%

6

Jo Magne Ingul. "Adolescent school absenteeism: modelling social and individual risk factors : Modelling risk factors for school absenteeism", Child and Adolescent Mental Health, 05/2012

Publication

<1%

7

lrias.kuleuven.be

Internet Source

<1 %

8

Submitted to Sheffield Hallam University

Student Paper

<1 %

9

arxiv.org

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On