

**KERTAS PELAPORAN
JAWATANKUASA PEMANDU
TEKNIKAL STATISTIK (JPTS)**

**KAJIAN PENENTUAN *VARIABLE OF INTEREST (VOI)*
UNTUK PENGIRAAN SAIZ SAMPEL: PENYIASATAN PERTUBUHAN**

Disediakan oleh:

Bahagian Metodologi dan Penyelidikan

Jabatan Perangkaan Malaysia

November 2016

Kertas Cadangan Dan Pelaporan Kajian Di Bawah Jawatankuasa Pemandu Teknikal Statistik

1. **Tajuk kajian:** Penentuan *Variable Of Interest* (VOI) Untuk Pengiraan Saiz Sampel - Penyiasatan Pertubuhan

1.1 Teras Strategik 1: Penghasilan produk dan perkhidmatan statistik yang berintegriti dan *reliable*.

1.2 Strategi 2: Meningkatkan penghasilan produk dan memperluaskan perkhidmatan statistik.

2. **Latar belakang**

Kajian *Variable Of Interest* (VOI) ini menggunakan peringkat analisis 5 digit *Malaysia Standard Industrial Classification* (MSIC) yang akan melibatkan 246 bilangan MSIC. VOI yang digunakan dalam analisis ini ialah:

- i. Hasil;
- ii. Pekerja; dan
- iii. Output.

2.1 **Pernyataan masalah (*Problem statement*)**

Terdapat dua pernyataan masalah yang dikenal pasti iaitu:

- i. Pelbagai indikator statistik dihasilkan daripada penyiasatan ekonomi; dan
- ii. Pengiraan saiz sampel bagi penyiasatan ekonomi menggunakan satu VOI sahaja.

3. **Objektif kajian**

Objektif kajian VOI ini dilaksanakan adalah:

- i. Untuk membuat perbandingan saiz sampel yang dikira menggunakan VOI:
 - a. Hasil;
 - b. Pekerja; dan

- c. Output.
- ii. Untuk mengenal pasti VOI yang menghasilkan saiz sampel optimum bagi penyiasatan pertubuhan.

4. Pendekatan metodologi

Metodologi kajian yang digunakan dalam kajian VOI ini ialah:

- i. Formula: *Estimation of population total*;
- ii. Penstrataan: Berdasarkan kategori Perusahaan Kecil Sederhana (PKS); dan
- iii. Dua rekabentuk persampelan.

4.1 Sumber data

Menggunakan data Banci Ekonomi 2011 (tahun rujukan 2010) bagi sektor pembuatan.

4.2 Pengiraan saiz sampel menggunakan formula berasaskan statistik jumlah

William G. Cochran (2011). Professor of Statistics, Emeritus, Harvard University. Sampling Techniques (third edition).

$$n = \frac{(\sum N_h \sigma_h)^2}{V + \sum N_h \sigma_h^2} \quad \text{dengan} \quad V = d \times \tau$$

$h = 1, 2, \dots, 246$

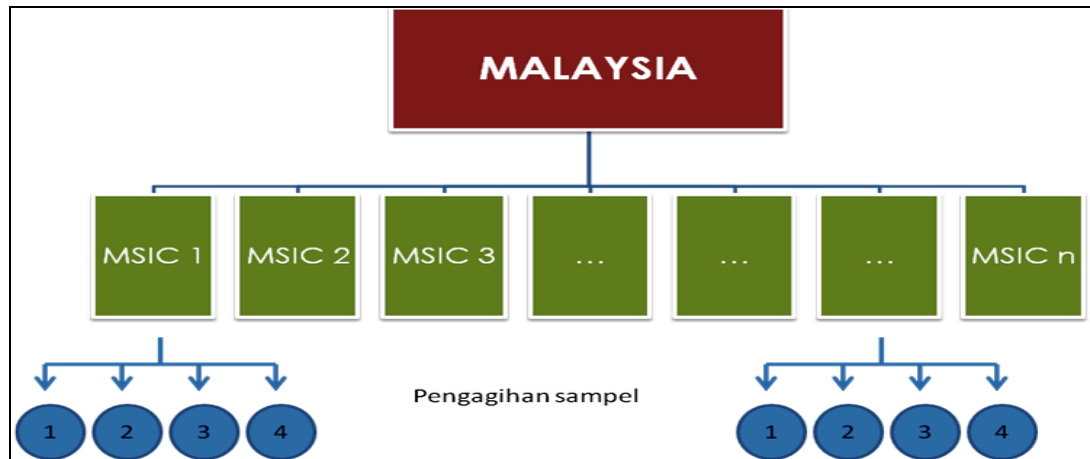
- N_h Populasi
- σ_h^2 Sisihan piawai bagi populasi
- d Ralat yang disasarkan
- σ_h Varians bagi populasi
- τ Jumlah nilai VOI
- h MSIC

4.3 Pembentukan substrata

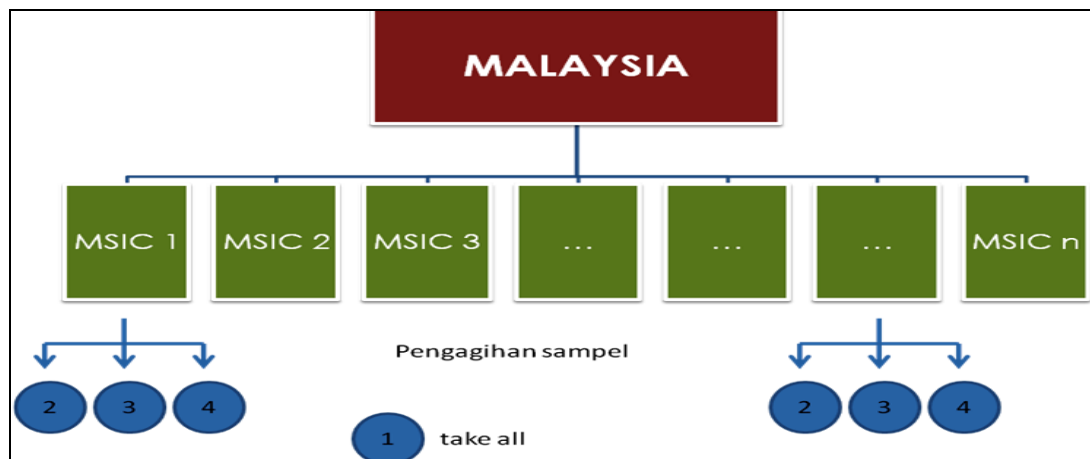
Varians nilai hasil/pekerja/output bagi setiap MSIC 5 digit adalah besar (*heterogenous*). Oleh itu, bagi setiap MSIC 5 digit data akan dikelaskan

kepada substrata 1 hingga 4 berdasarkan kategori Perusahaan Kecil dan Sederhana (PKS).

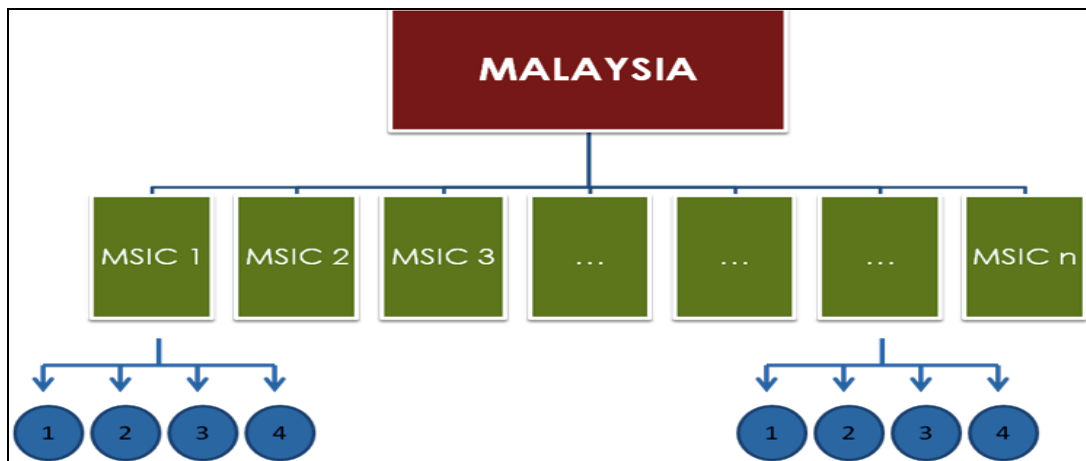
Rekabentuk 1: Pengiraan saiz sampel di peringkat MSIC



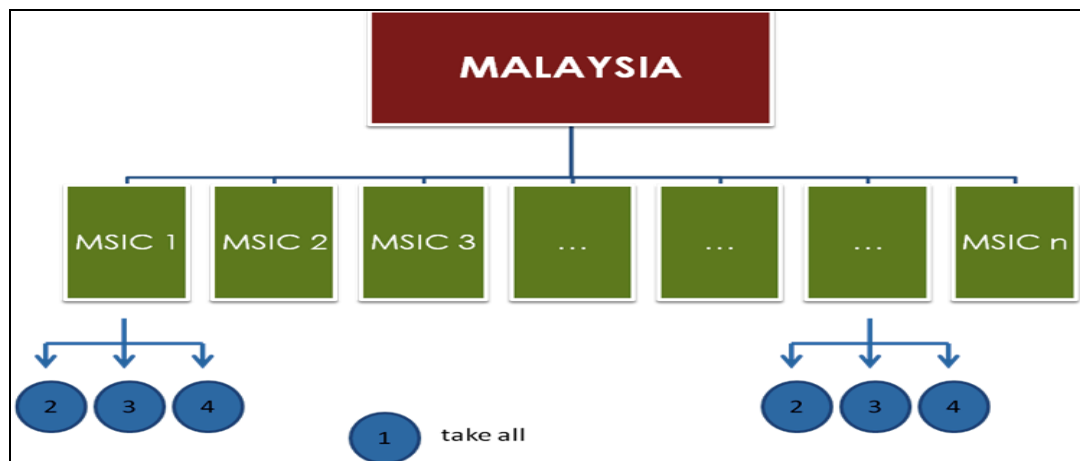
Reka bentuk 1a (strata 1 take all): Pengiraan saiz sampel di peringkat MSIC. Pertubuhan yang dikenal pasti mempunyai sumbangan yang signifikan, take all dan selainnya take some.



Rekabentuk 2: Pengiraan saiz sampel di peringkat MSIC dan substrata



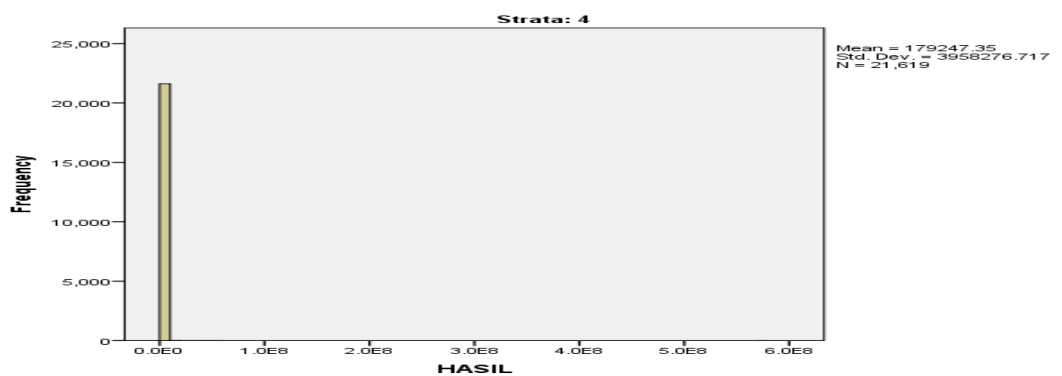
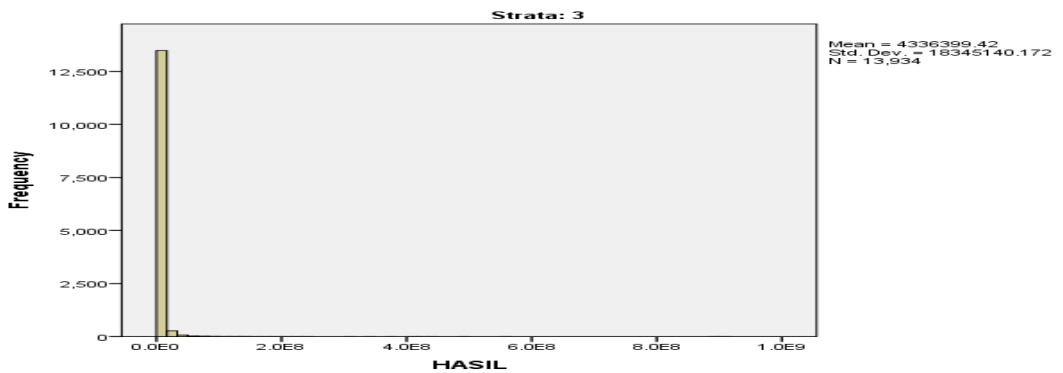
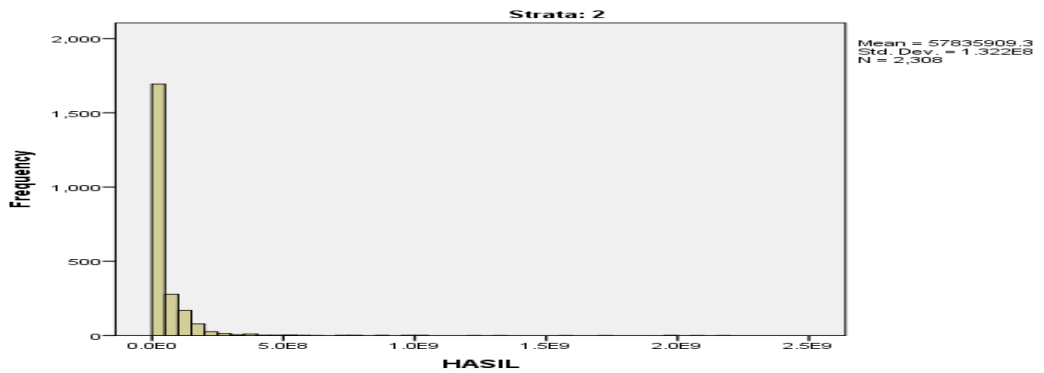
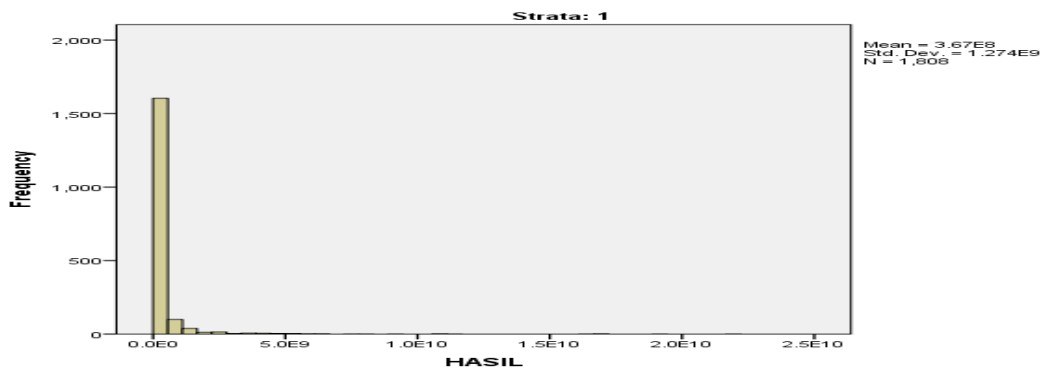
Rekabentuk 2a (strata 1 take all): Pengiraan saiz sampel di peringkat MSIC dan substrata. Pertubuhan yang dikenalpasti mempunyai sumbangan yang signifikan, *take all* dan selainnya *take some*.



4.4 Data profiling

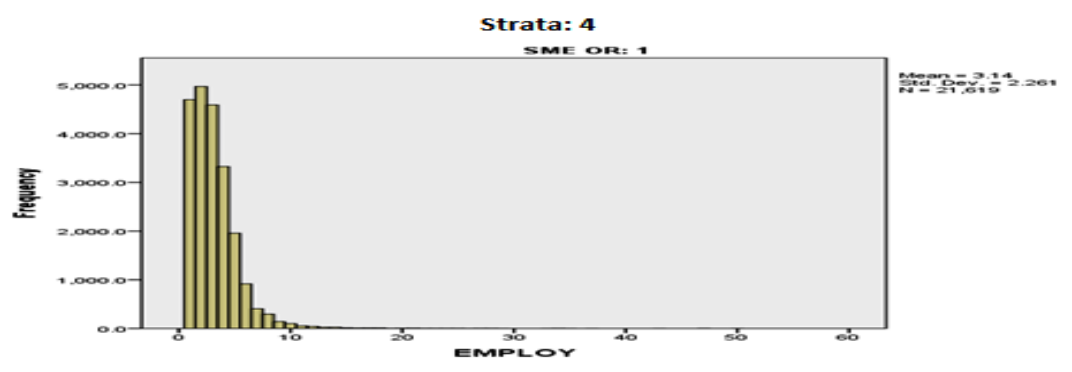
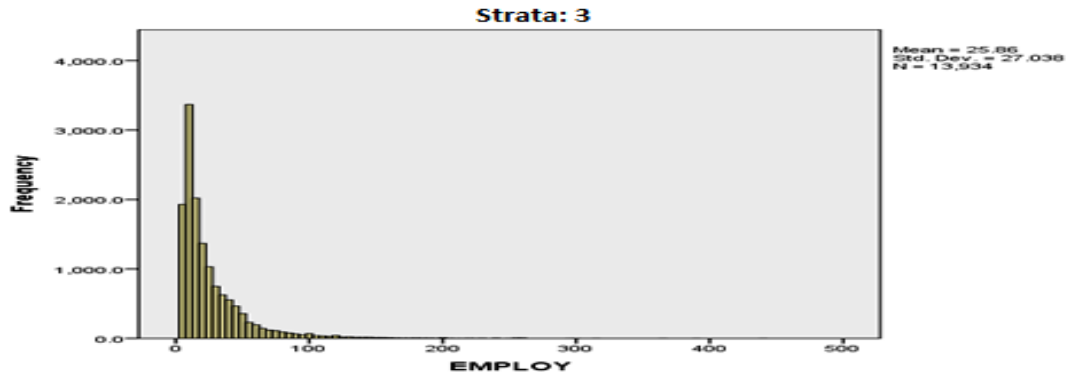
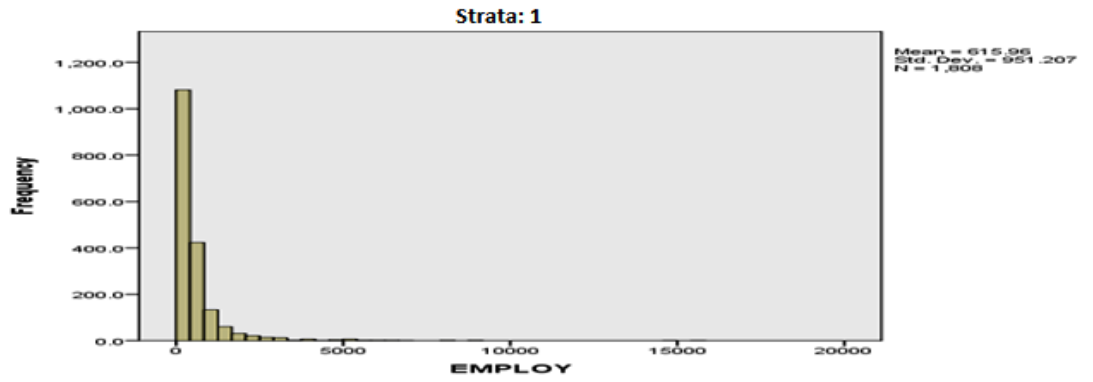
Hasil (RM)

<i>Mean</i>	21,699,279
<i>Median</i>	277,153
<i>Mode</i>	18,000
<i>Std. Deviation</i>	284,404,353
<i>Minimum</i>	332 (n=1 ; 0.00%)
<i>Maximum</i>	21,985,924,606



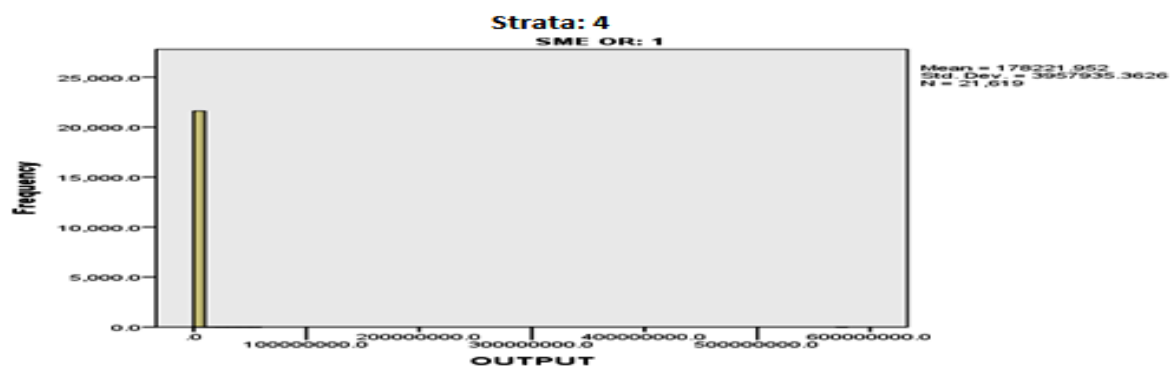
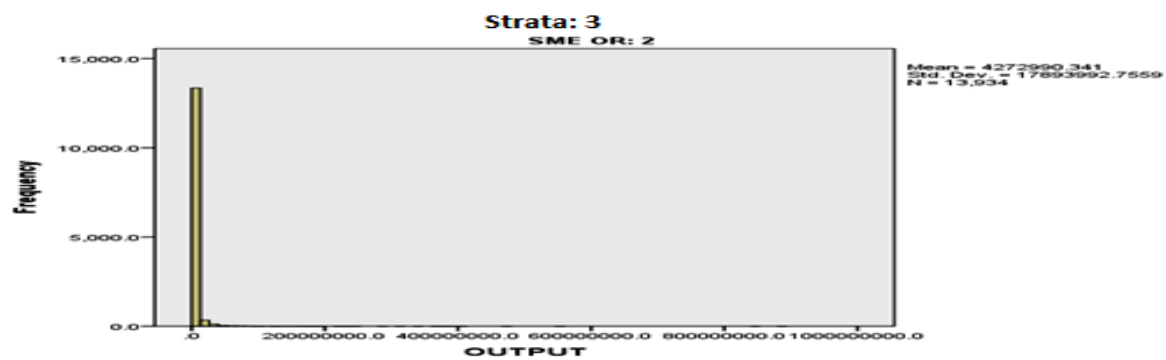
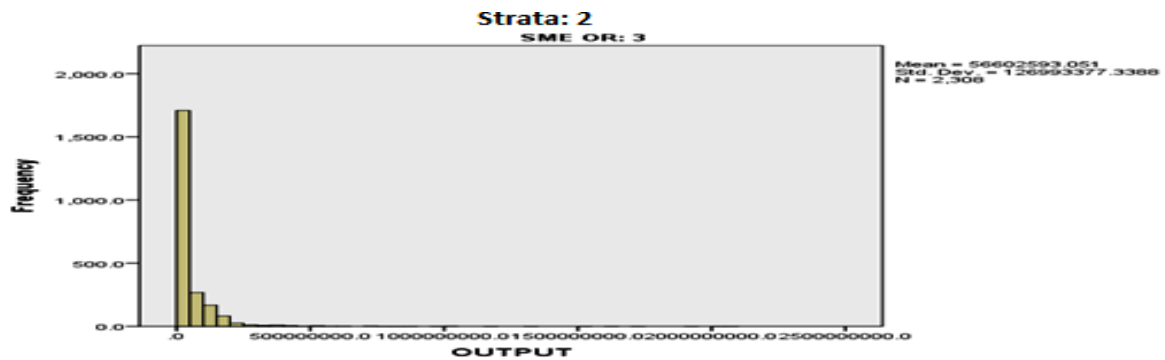
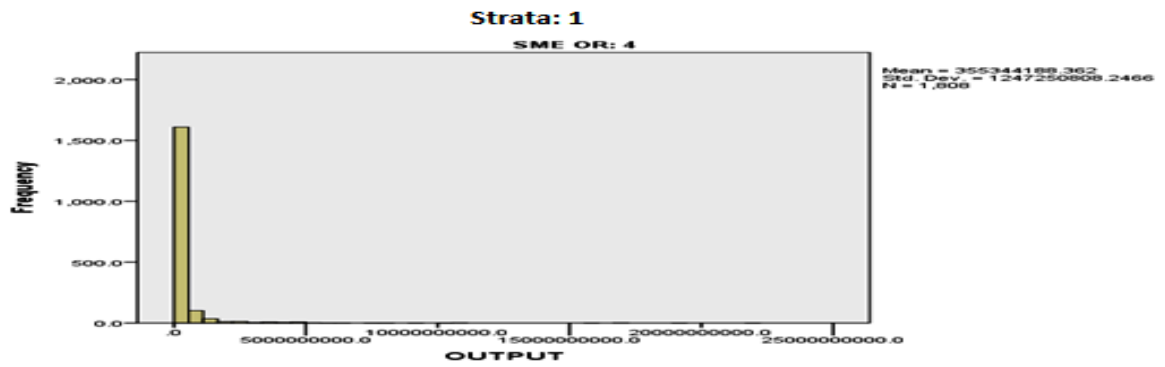
Pekerja

<i>Mean</i>	46
<i>Median</i>	5
<i>Mode</i>	2
<i>Std. Deviation</i>	241
<i>Minimum</i>	1 (n=4,699 ; 11.8%)
<i>Maximum</i>	15,676



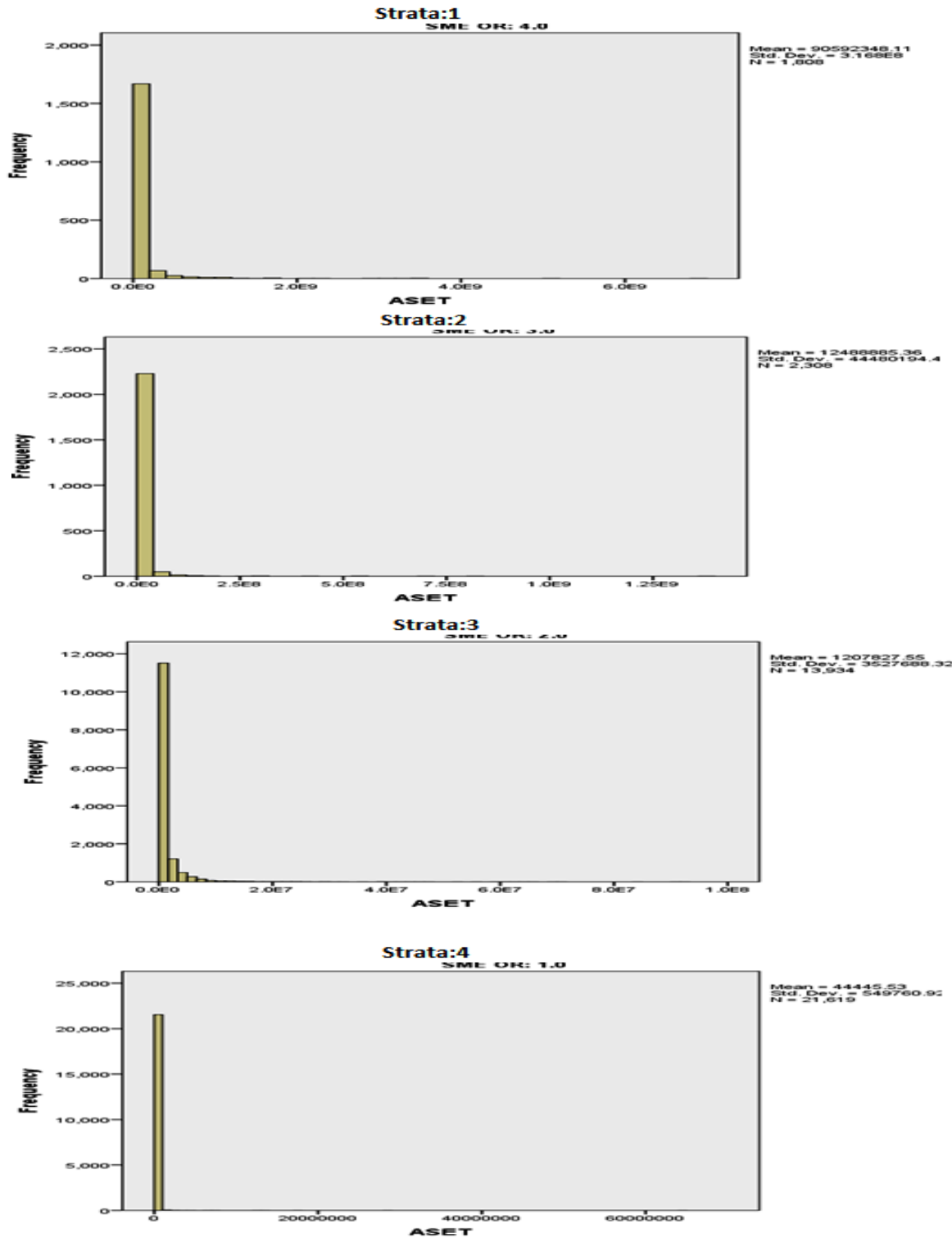
Output (RM)

<i>Mean</i>	21,086,841
<i>Median</i>	275,429
<i>Mode</i>	18,000
<i>Std. Deviation</i>	278,255,667
<i>Minimum</i>	332 (n=1 ; 0.00%)
<i>Maximum</i>	21,932,011,533



Aset (RM)

Mean	5,304,042
Median	34,711
Mode	1
Std. Deviation	71,036,951
Minimum	1 (n=1,779 ; 4.5%)
Maximum	6,830,609,000



5. Hasil dan rumusan

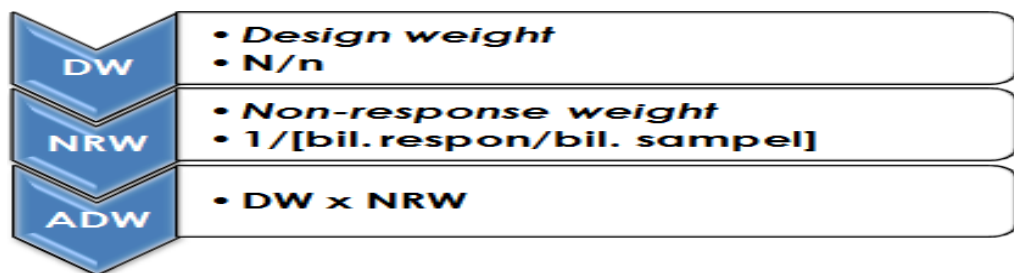
5.1 Penemuan kajian

Saiz sampel optimum

REKABENTUK	RALAT	HASIL	PEKERJA	OUTPUT
1. MSIC 5D	5% 10%	1,477 759	2,032 705	1,476 762
1a. MSIC 5D (strata 1 take all)	5% 10%	2,101 2,059	2,114 2,060	2,101 2,059
2. MSIC 5D dan Substrata	5% 10%	1,015 896	977 870	1,011 896
2a. MSIC 5D dan Substrata (strata 1 take all)	5% 10%	2,471 2,467	2,472 2,467	2,471 2,467

5.2 Analisis data berpemberat

i. Pengiraan *weight*



ii. *Actual vs. Estimated*

Berdasarkan rekabentuk 2a

Saiz sampel: 2,472

Item	Actual (BE 2011)	Estimated	% Diff.
Hasil (RM)	860,788,682,092	925,451,076,054	7.51
Pekerja	1,812,360	1,766,337	(2.54)
Output (RM)	836,493,905,119	898,742,199,202	7.44
Population Size	39,669	39,669	-

5.3 Rumusan

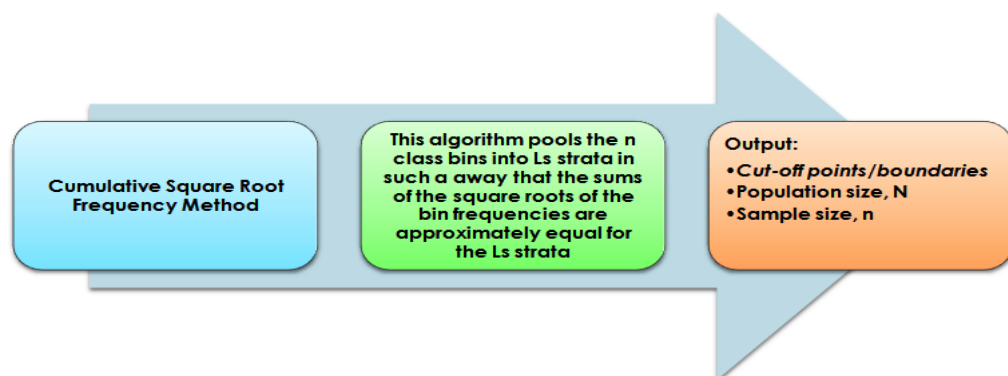
- i. Saiz sampel optimum dipengaruhi oleh:
 - a. Kaedah yang digunakan;
 - b. Nilai ralat yang disasarkan; dan
 - c. Varian
- ii. Saiz sampel optimum dipengaruhi oleh **variasi yang wujud dalam data** (VOI). Oleh itu, penetapan VOI bagi menghasilkan saiz sampel optimum sangat bergantung kepada kualiti data penyiasatan/banci.
- iii. Saiz sampel dikira berdasarkan semua VOI dalam penyiasatan dan saiz sampel yang paling optimum dipilih.

5.4 *Moving forward*

Penambahbaikan kaedah penstrataan dengan menggunakan:

- i. *Sophie Baillargeon and Louis-Paul Rivest (2011). The Construction of Stratified Designs in R with the Package Stratification, Survey Methodology, Statistics Canada.*

Pembentukan Strata: Stratification Package dalam Perisian R

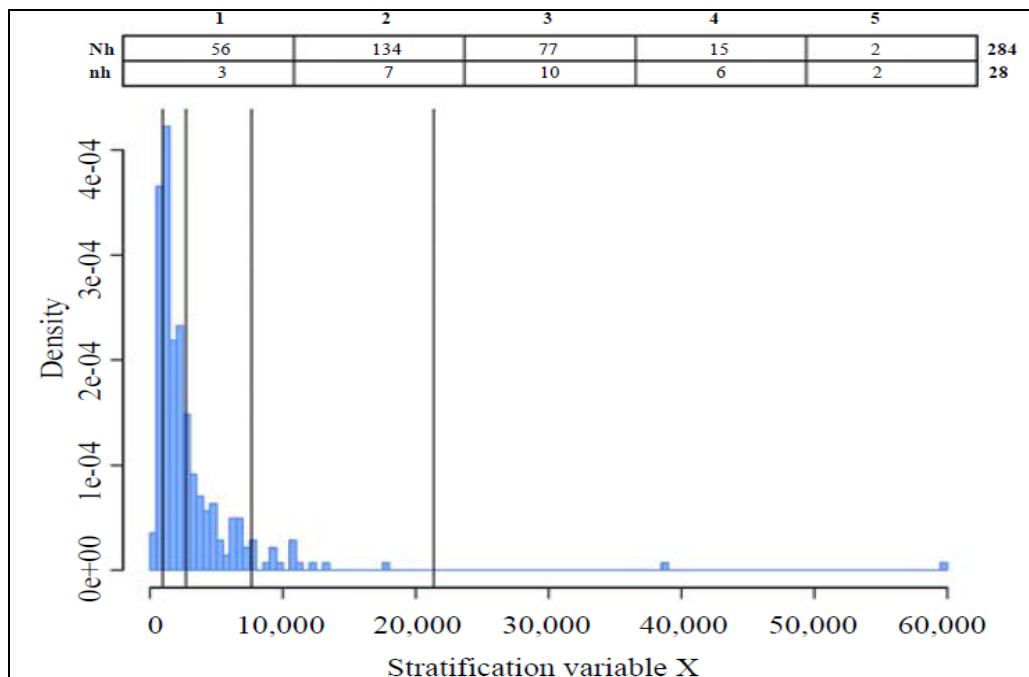


Contoh: Cumulative Square Root Frequency Method

Variable: Output

	<i>Lower</i>	<i>Upper</i>	<i>Frequency</i>	<i>Sqrtf</i>	<i>CumSqrtf</i>	<i>Stratum</i>
1	0	5	3464	58.86	58.86	1
2	5	10	2516	50.16	109.02	2
3	10	15	2157	46.44	155.46	2
4	15	20	1581	39.76	195.22	3
5	20	25	1142	33.79	229.01	3
6	25	30	746	27.31	256.33	4
7	30	35	512	22.63	278.95	4
8	35	40	376	19.39	298.35	4
9	40	45	265	16.28	314.62	4
10	45	50	207	14.39	329.01	5
11	50	55	126	11.22	340.24	5
12	55	60	107	10.34	350.58	5
13	60	65	82	9.06	359.64	5
14	65	70	50	7.07	366.71	5
15	70	75	39	6.24	372.95	5
16	75	80	25	5.00	377.95	5
17	80	85	16	4.00	381.95	5
18	85	90	19	4.36	386.31	5
19	90	95	2	1.41	387.73	5
20	95	100	3	1.73	389.46	5
Cutpoints	77.89	155.78	233.67	311.57		

Graphical Representation of the Stratification



Saiz sampel optimum

REKABENTUK	Substrata	RALAT	HASIL	PEKERJA	OUTPUT
2.MSIC 5D dan Substrata	PKS	5% 10%	94 77	186 71	91 75
	CSRF Method	5% 10%	226 135	253 132	226 133
2a.MSIC 5D dan Substrata (strata 1 take all)	PKS	5% 10%	268 251	355 245	265 249
	CSRF Method	5% 10%	252 235	313 280	251 231