Uji Kesamaan Distribusi Data *Zero Inflated* pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia

Test of Homogeneity Distribution in Cases of Zero Inflated Problem Applied to Claims Frequency Data of Motor Vehicle Insurance in Indonesia

[1]Atik Rohayati, [2]Suliadi, [3]Anneke Iswani Achmad
[1,2,3]Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Jl. Tamansari No.1 Bandung 40116
email: [1]atik.rohayati27@gmail.com, [2]suliadi@gmail.com

Abstract. This paper discusseshomogeneity test ofdistribution ofzero inflated datausing percentiles profile. Zero inflated is data which the proportion of zero value far greater than the proportion of the others. There are some cases where researchers want to compare *k* population groups, with zero inflated problem. The researchers want to compare whether the distributionsof allare the same. In this paper we use chi-square test to compare equality of distributions of claim frequency data of motor vehicle insurance in Indonesia that can handle zero inflated problem.The results showed thatthe five classes of claims frequency data have different distributions, thus in the analysis of claim frequency data, they should be splitted among classes of motor vehilcle in Indonesia.

Keywords:zero inflated, percentiles profile, claims frequency, Chi-Square.

Abstrak. Makalah ini membahas pengujian kesamaan distribusi data *zero infalted* menggunakan profil persentil. *Zero inflated* adalah data dimana proporsi nilai nol jauh lebih besar daripada proporsi bukan nilai nol. Terdapat beberapa kasus dimana peneliti ingin membandingkan *k* kelompok populasi, yang masing-masing populasi itu terdapat *zero inflated*. Pada kasus perbandingan *k* kelompok populasi peneliti ingin membandingkan apakah distribusi antara kelompok satu dengan kelompok yang lain adalah sama. Dalam makalah ini kami menggunakan uji *chi-square* untuk membandingkan kesamaan distribusi data frekuensi klaim asuransi kendaraan bermotor di Indonesia yang dapat menangani masalah *zero inflated.*Hasil penelitian menunjukkan bahwa lima kelas dari data klaim frekuensi memiliki distribusi yang berbeda, dengan demikian dalam analisis data frekuensi klaim, mereka harus dipisahkan antara masing-masing kelas asuransi kendaraan bermotor di Indonesia.

Kata Kunci:*zero inflated*, profil persentil, frekuensi klaim, *Chi-Square*.

A. Pendahuluan

Pada data pengamatan sering kali ditemukan dimana proporsi nilai nol jauh lebih besar daripada proporsi bukan nilai nol, kasus ini disebut dengan data *zero inflated*. Distribusi *zero inflated* muncul dalam berbagai kondisi. Sebagai contoh, beberapa prosedur laboratorium yang mungkin tidak dapat mendeteksi nilai yang berada di bawah ambang batas tertentu dan secara umum dicatat sebagai nol. Namun distribusi *zero inflated* umumnya terjadi pada data asuransi khususnya data frekuensi klaim.

Secara umum klaim asuransi adalah tuntutan dari pihak tertanggung sehubungan dengan adanya kontrak perjanjian antara asuransi dengan pihak tertanggung. Dimana masing-masing pihak mengikatkan diri untuk menjamin pembayaran ganti rugi oleh penanggung. Hal ini terjadi jika pembayaran premi asuransi telah dilakukan oleh pihak tertanggung. Dan dilakukan pembayaran ganti rugi ketika terjadi musibah yang diderita oleh pihak tertanggung.Data asuransi terdiri dari data besar klaim dan frekuensi klaim. Menurut Klugman dkk. (2004), frekuensi klaim adalah banyaknya klaim yang dilakukan oleh seorang pemegang polis selama masa asuransinya.

ada beberapa kasus dimana peneliti ingin membandingkan *k* kelompok sampel atau populasi yang masing-masing sampel itu terdapat *zero inflated.* Pada kasus perbandingan *k* kelompok sampel peneliti ingin membandingkan apakah distribusi antara kelompok satu dengan kelompok yang lain adalah sama. Lachenbruch (2001) mengusulkan untuk pengujian kesamaan distribusi *k* kelompok sampel untuk kasus *zero inflated*, dengan memisahkan pengujian antara kategori nol dan kategori lainnya.

Sedangkan, Johnson dkk (2015) mengajukan sebuah metode sederhana untuk menguji apakah *k* kelompok sampel yang saling bebas berasal dari distribusi yang sama atau mempunyai distribusi yang berbeda dengan menggunakan uji *chi-square.* Hal ini diaplikasikan Johnson pada data *zero inflated*. Metode ini mempunyai sifat-sifat yang baik untuk sampel yang berukuran besar. Selain itu, pengujian ini juga bisa digunakan ketika berhadapan dengan data yang tidak simetris atau dalam kasus dimana distribusinya tidak diketahui. Metode ini juga dapat digunakan untuk data kontinu atau data diskrit, selama sampelnya berukuran besar. Untuk membandingkan kesamaan dua populasi atau lebih metode pengujian ini lebih fleksibel dibandingkan dengan uji median, karena metode ini menggunakan profil persentil sehingga informasi yang dipergunakan lebih banyak dibandingkan dengan uji median atau uji rata-rata. Pada skripsi ini kami ingin menerapkan metode yang disarankan oleh Johnson untuk menguji apakah distribusi frekuensi klaim asuransi untuk berbagai kategori kendaraan bermotor di Indonesia sama atau tidak.

Dalam makalah ini kami menguji kesamaan distribusi beberapa kelompok asuransi kendaraan bermotor di Indonesia pada tahun 2011, dengan menggunakan metode yang disarankan oleh Johnson dkk. pada tahun 2015. Metode ini dapat digunakan untuk menguji kesamaan distribusi ketika ada masalah *zero inflated* pada data.

B. Landasan Teori

Asuransi Kendaraan Bermotor

Asuransi terbagi menjadi dua jenis yaitu asuransi jiwa dan kerugian. Salah satu contoh asuransi kerugian adalah asuransi kendaraan bermotor. Asuransi kendaraan bermotor adalah produk asuransi kerugian yang melindungi tertanggung dari risiko kerugian yang mungkin timbul sehubungan dengan kepemilikan dan pemakaian kendaraan bermotor.Ada dua jenis perlindungan untuk asuransi kendaraan bermotor, yaitu *Total Loss Only* (TLO) dan *Comprehensive* (Komprehensif).

Dalam produk asuransi TLO, jenis klaim yang diajukan adalah *total loss*.Dalam produk asuransi *Comprehensive*, jenis klaim yang diajukan ada dua kemungkinan yaitu, *total loss* dan *partial loss.* Perlu diketahui bahwa klaim *total loss* hanya dapat diajukan sekali oleh tertanggung ke penanggung. Sedangkan klaim *partial loss* dapat diajukan lebih dari sekali oleh tertanggung ke penanggung.

Uji Kesamaan Beberapa Distribusi Berdasarkan Distribusi *Chi-Square*

Secara garis besar ilmu statistika dibagi menjadi dua bagian yaitu statistika parametrik dan statistika nonparametrik. Statistika parametrik adalah ilmu statistika yang digunakan untuk data-data yang diasumsikan berasal dari populasi yang berdistribusi tertentu. Statistika nonparametrik disebut juga statistika bebas distribusi. Metode statistika nonparametrik merupakan metode statistika yang dapat digunakan dengan mengabaikan asumsi-asumsi yang melandasi penggunaan metode statistika parametrik. Salah satu metode statistika nonparametrik adalah uji kesamaan beberapa distribusi. Uji-uji tersebut adalah uji Birnbaum-Hall dan uji *k*-sampel Smirnov (Conover, 1980). Namun kedua uji ini tidak dapat digunakan karena sampelnya yang terbatas dan data yang digunakan adalah berdistribusi kontinu. Menurut Daniel (1989) untuk uji kesamaan distribusi *k* kelompok sampel menggunakan uji *chi-square.*

Daniel (1989) menjelaskan pengujian kesamaan distribusi*k* kelompok sampel menggunakan uji *chi-square*. Uji *chi-square* adalah teknik analisis yang digunakan untuk menentukan perbedaan frekuensi observasi (O_i) dengan frekuensi ekspektasi atau frekuensi harapan (E_i) untuk suatu kategori tertentu. Uji ini dapat dilakukan pada data diskrit atau frekuensi.

Data untuk analisis dengan cara ini disusun dalam tabel kontingensi seperti pada Tabel 2.2.

Tabel 2.2 Tabel Kontingensi *k* x *p*

		Kolom					Total
		1	2	3	⋯	p	
Baris	1	O_{11}	O_{12}	O_{13}	⋯	O_{1p}	$T_{1.}$
	2	O_{21}	O_{22}	O_{23}	⋯	O_{2p}	$T_{2.}$
	⋮	⋮	⋮	⋮	⋱	⋮	⋮
	K	O_{k1}	O_{k2}	O_{k3}	⋯	O_{kp}	$T_{k.}$
	Total	$T_{.1}$	$T_{.2}$	$T_{.3}$	⋯	$T_{.p}$	$T_{..}$

dimana:

O_{ij} : frekuensi yang nilai observasi yang terdapat dalam baris ke-*i* dan kolom ke-*j*

$T_{i.}$: frekuensi keseluruhan pada baris ke-*i*, dimana $i = 1,..,k$

$T_{.j}$: frekuensi keseluruhan pada kolom ke-*j*, dimana $j = 1,..,p$

$T_{..}$: frekuensi keseluruhan O_{ij}

Salah satu metode statistika nonparametrik adalah uji kesamaan beberapa distribusi untuk *k* sampel *independent*. Jika kita memiliki data *k* kelompok sampel dan masing – masing sampelnya terdapat beberapa *p* kategori. Untuk pengujian kesamaan beberapa distribusi ini menggunakan uji *chi-square* dengan perumusan hipotesisnya sebagai berikut:

H_0: Tidak ada perbedaan distribusi diantara *k* kelompok baris.

H_1: Minimal ada dua kelompok baris yang distribusinya berbeda.

Seperti halnya dengan analisa untuk tabel kontingensi statistik yang digunakan adalah *chi-square*:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{p} \left\{ \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \right\} \qquad \dots (2.1)$$

dengan penjumlahan dilakukan terhadap semua (k x p) sel.

Untuk dapat menghitung statistik uji di atas, maka terlebih dahulu kita perlu menghitung frekuensi harapan bagi setiap sel dalam Tabel 2.2 di bawah asumsi bahwa H_0 benar. Rumus umum untuk mendapatkan frekuensi harapan ini adalah :

$$E_{ij} = \frac{T_{.j} \times T_{i.}}{T_{..}} \qquad \dots (2.2)$$

Adapun kriteria uji dalam pengujian hipotesis di atas adalah bila $\chi^2 > \chi^2_{(1-\alpha)}$ dengan derajat bebas $\nu = (k-1)(p-1)$, maka hipotesis H_0 ditolak artinya minimal ada dua sampel yang distribusinya berbeda, bila dalam hal lainnya H_0 diterima. Namun, jika setiap sampel atau populasi pada data mempunyai distribusi yang sama maka pemodelan dapat digabungkan.

Daniel (1989), menyatakan nilai ekspektasi tidak boleh kurang dari 5. Hal ini diperbolehkan jika nilai ekspektasi yang kurang dari 5 maksimal 20% dari jumlah sel. Sebagai contoh, tabel kontingensi 4×4 terdapat nilai ekspektasi yang kurang dari 5 sebanyak 3 sel, penggunaan *chi-square* dapat dilanjutkan. Namun jika nilai ekspektasi yang kurang dari 5 lebih dari 20% dari banyaknya sel, maka harus dilakukan penggabungan sel atau dilakukan koreksi Yates.

Uji Kesamaan Populasi pada Kasus *Zero Inflated*

Data *zero inflated* biasanya ditemukan dalam beberapa kasus misalkan pada kasus asuransi. Kadang-kadang peneliti ingin membandingkan apakah beberapa kelompok sampel memiliki distribusi yang sama atau tidak. Untuk banyaknya data nol pada data frekuensi, beberapa metode telah diusulkan untuk menguji kesamaan data *zero inflated*. Tse dkk. (2009) dan Yuen dkk. (2015) mengusulkan untuk menggunakan metode rasio likelihood, sedangkan Bedrick dan Hossain (2013) mengusulkan untuk menggunakan uji eksak. Namun demikian hal itu menimbulkan masalah karena metode-metode tersebut mengasumsikan datanya berdistribusi Poisson. Sedangkan untuk data yang mengandung proporsi nolnya jauh lebih banyak daripada selain nol, dan diasumsikan mengikuti distribusi parametrik tertentu, uji standar seperti uji rasio likelihood dan uji Wald tidak dapat bekerja dengan baik jika bentuk distribusinya tidak diketahui, sebab akan terjadi inflasi pada kesalahan tipe satu. Sedangkan Hallstrom (2010) menggunakan uji Wilcoxon terpangkas, dimana data yang dipangkasnya adalah data yang nilainya nol.

Johnson dkk (2015) mengajukan metode untuk menguji kesamaan distribusi data *zero inflated*. Metode ini mempunyai sifat-sifat yang baik untuk sampel yang berukuran besar. Selain itu, pengujian ini juga bisa digunakan ketika berhadapan dengan data yang tidak simetris atau dalam kasus dimana distribusinya tidak diketahui. Metode ini juga dapat digunakan untuk data kontinu atau data diskrit, selama sampelnya berukuran besar. Untuk membandingkan kesamaan dua populasi atau lebih metode pengujian ini lebih fleksibel dibandingkan dengan uji median, karena metode ini menggunakan profil persentil sehingga informasi yang dipergunakan lebih banyak dibandingkan dengan uji median atau uji rata-rata.

C. Hasil Penelitian dan Pembahasan

Tabel Kontingensi

Data yang digunakan adalah data sekunder hasil pencatatan yang diperoleh dari Kementerian Keuangan Republik Indonesia pada tahun 2012. Data tersebut berisi tentang pemegang polis asuransi kendaraan bermotor yang melakukan klaim terhadap perusahaan asuransi kerugian yang menaunginya. Data yang akan dipakai untuk keperluan aplikasi adalah data pemegang polis asuransi kendaraan bermotor yang terdiri dari 5 kelompok sesuai dengan harga pertanggungan, dimana klaim yang diajukannya adalah *Partial Loss* dan tersaji pada Tabel 3.1.

Kemudian sebelum dilakukan pengujian kesamaan distribusi pada data frekuensi klaim pemegang polis asuransi kendaraan bermotor kelompok 1, 2, 3, 4, dan 5 di Indonesia, terlebih dahulu membentuk tabel kontingensi $5 \times (p+2)$. Langkah pertamamemisalkan profil persentil pada setiap kelompok populasi adalah $Q_i = \left(Q_{i1}, Q_{i2}, \dots, Q_{ip}\right)$, dimana $i = 1, 2, 3, 4, 5$.

Tabel 3.1 Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia Kelompok 1 s/d 5 Tahun 2011

Frekuensi Klaim	Kelompok				
	1	2	3	4	5
0	299.542	180.872	30.831	5.977	6.688
1	66.909	55.727	6.848	985	397
2	23.613	19.264	2.036	342	95
3	8.607	6.506	632	117	28
4	3.453	2.406	223	37	11
5	1.401	896	61	14	1
6	552	339	29	4	-
7	251	143	13	3	-
8	132	66	3	2	-
9	58	28	2	-	-
10	33	16	1	-	-
11	20	11	-	-	-
12	12	3	3	-	-
13	-	3	1	-	-
14	3	-	-	-	-
15	-	2	-	-	-
16	-	-	-	-	-
17	2	-	-	-	-
21	1	-	-	-	-

Langkah selanjutnya adalahmemisahkan nilai nol dari setiap k populasi dan dijadikan sebagai kategori ke-0 atau (O_{i0}).Hasil dari memisahkan kategori nol ini disajikan pada Tabel 3.2.

Tabel 3.2 Banyaknya Pemegang Polis yang Tidak Mengajukan Klaim (O_{h0})

Kelompok Populasi ke-	Banyaknya Pemegang Polis
1	299.542
2	180.872
3	30.831
4	5.977
5	6.688
Total	523.910

Setelah itu menggabungkan dan mengurutkan kategori selain nol dari kecil ke besar yang ada di ke-5 kelompok populasi.Kemudian menghitung penaksir persentil profil pada populasi gabungan yang dinyatakan sebagai$q = \left(q_1, q_2, \dots, q_p\right)$dengan nilai p = 5 maka $q = (q_1, q_2, q_3, q_4, q_5)$. Untuk mendapatkan nilai q maka dapat dilihat pada Tabel 3.3.

Tabel 3.3 PersentaseBanyaknya Pemegang Polis yang Mengajukan Klaim

#Klaim	**Frekuensi**	**%**	**% Kumulatif**
1	130.866	64,67469	64,67469
2	45.350	22,41222	87,08691
3	15.890	7,852924	94,93983
4	6.130	3,029479	97,96931
5	2.373	1,17275	99,14206
6	924	0,456646	99,59871
7	410	0,202624	99,80133
8	203	0,100324	99,90165
9	88	0,04349	99,94514
10	50	0,02471	99,96985
11	31	0,01532	99,98517
12	18	0,008896	99,99407
13	4	0,001977	99,99605
14	3	0,001483	99,99753
15	2	0,000988	99,99852
17	2	0,000988	99,99951
21	1	0,000494	100
Total	202.345	100	

Dari Tabel 3.3 dapat dilihat bahwa banyak klaim yang diajukan pemegang polis sebanyak satu mencapai 64%. Hal ini menunjukan bahwa terdapat perubahan banyak klaim yang diajukan terjadi pada persentil ke-60, oleh karena itu diambil persentil ke-50 untuk mendapatkan nilai persentil sama dengan 1. Untuk menentukan nilai persentil lainnya disajikan sebagai berikut:

q_1 = persentil ke-50 dengan nilai persentilnya 1
q_2 = persentil ke-65 dengan nilai persentilnya 2
q_3 = persentil ke-88 dengan nilai persentilnya 3
q_4 = persentil ke-95 dengan nilai persentilnya 4
q_5 = persentil ke-98 dengan nilai persentilnya 5

Setelah menentukan nilai persentilnya, langkah selanjutnya adalah membentuk tabel kontingensi $5 \times (p + 2)$ dengan menghitung pengamatan frekuensi setiap kelompok populasi ke-i untuk masing-masing kategori (bin), dimana $i = 1,2,3,4,5$ dan p = 5. Tabel kontingensi untuk kasus ini disajikan pada Tabel 3.4

Tabel 3.4 Tabel Kontingensi $5 \times (p+2)$

		Bin ke-							Total
		0	1	2	3	4	5	6	
Kelompok Populasi ke-	1	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	O_{16}	$T_{1.}$
	2	O_{20}	O_{21}	O_{22}	O_{23}	O_{24}	O_{25}	O_{26}	$T_{2.}$
	3	O_{30}	O_{31}	O_{32}	O_{33}	O_{34}	O_{35}	O_{36}	$T_{3.}$
	4	O_{40}	O_{41}	O_{42}	O_{43}	O_{44}	O_{45}	O_{46}	$T_{4.}$
	5	O_{50}	O_{51}	O_{52}	O_{53}	O_{54}	O_{55}	O_{56}	$T_{5.}$
	Total	$T_{.0}$	$T_{.1}$	$T_{.2}$	$T_{.3}$	$T_{.4}$	$T_{.5}$	$T_{.5}$	$T_{..}$

Untuk menentukan O_{ij}pada kelompok populasi ke-1 dengan cara seperti berikut :

$$\text{bin}_1 : O_{11} = \{\text{banyak pengamatan populasi ke 1} \le q_1\},$$
$$\text{bin}_2 : O_{12} = \{q_1 < \text{banyak pengamatanpopulasi ke 1} \le q_2\},$$
$$\text{bin}_3 : O_{13} = \{q_2 < \text{banyak pengamatanpopulasi ke 1} \le q_3\},$$
$$\text{bin}_4 : O_{14} = \{q_3 < \text{banyak pengamatanpopulasi ke 1} \le q_4\},$$
$$\text{bin}_5 : O_{15} = \{q_4 < \text{banyak pengamatanpopulasi ke 1} \le q_5\},$$
$$\text{bin}_6 : O_{16} = \{\text{banyak pengamatan populasi ke 1} > q_6\}.$$

Maka hasil untuk O_{ij} pada kelompok populasi ke-1 tersaji pada Tabel 3.5.

Tabel 3.5 Hasil Perhitungan Nilai Bin pada Populasi ke-1

Kelompok Populasi ke-	Bin ke-					
	1	2	3	4	5	6
1	66.909	23.613	8.607	3.453	1.401	1.064

Untuk menentukan O_{ij} pada kelompok populasi ke-2, 3, 4, dan 5 dilakukan dengan perhitungan seperti diatas. Maka hasil keseluruhan dari perhitungan ini disajikan pada Tabel 3.6.

Tabel 3.6 Hasil Perhitungan Nilai Bin pada Populasi ke 1, 2, 3, 4, dan 5

Kelompok Populasi ke-	Bin ke-					
	1	2	3	4	5	6
1	66.909	23.613	8.607	3.453	1.401	1.064
2	55.727	19.264	6.506	2.406	896	611
3	6.848	2.036	632	223	61	52
4	985	342	117	37	14	9
5	397	95	28	11	1	0

Setelah melakukan perhitungan untuk menentukan nilai Bin selanjutnya menggabungkan O_{i0} dari langkah sebelumnya, sehingga terbentuk tabel kontingensi $5 \times (p+2)$ seperti pada Tabel 3.7.

Tabel 3.7 Tabel Kontingensi $5 \times (p+1)+1$

		Bin ke-							Total
		0	1	2	3	4	5	6	
Kelompok Populasi ke-	1	299.542	66.909	23.613	8.607	3.453	1.401	1.064	404.589
	2	180.872	55.727	19.264	6.506	2.406	896	611	266.282
	3	30.831	6.848	2.036	632	223	61	52	40.683
	4	5.977	985	342	117	37	14	9	7.481
	5	6.688	397	95	28	11	1	0	7.220
	Total	523.910	130.866	45.350	15.890	6.130	2.373	1.736	726.255

Uji Kesamaan Distribusi

Uji *Chi-square*

Dalam bagian ini akan dilakukan pengujian kesamaan distribusi pada data frekuensi klaim pemegang polis asuransi kendaraan bermotor kategori 1, 2, 3, 4, dan 5 di Indonesia menggunakan uji *chi-square*. Hipotesis untuk pengujian tersebut adalah:

H_0 : $Q_1 = Q_2 = Q_3 = Q_4 = Q_5$; profil persentil kelompok populasi sama.

H_1 : Minimal ada dua kelompok populasi yang profil persentilnya tidak sama.

Melakukan pengujian kesamaan distribusi profil persentil dengan statistik uji *chi-square* sebagaimana menggunakan Persamaan 2.1. Namun terlebih dahulu menentukan ekspektasi frekuensi setiap observasi dan dihitung menggunakan Persamaan 2.2. Berikut hasil perhitungan dari Persamaan 2.2.

$$E_{10} = \frac{T_{.0} \times T_{1.}}{T_{..}} = \frac{523.910 \times 404.589}{726.255} = 291.864{,}735$$

… dst

Hasil perhitungan ekspektasi frekuensi setiap observasi disajikan pada Tabel 3.8

Tabel 3.8 Hasil Perhitungan Nilai Ekspektasi

		Bin ke-						
		0	1	2	3	4	5	6
Kelompok Populasi ke-	1	291.864,735	72.904,068	25.264,007	8.852,151	3.414,958	1.321,973	967,107
	2	192.092,037	47.982,128	16.627,615	5.826,082	2.247,570	870,062	636,506
	3	29.348,136	7.330,788	2.540,394	890,118	343,387	132,930	97,246
	4	5.396,687	1.348,023	467,141	163,680	63,144	24,444	17,882
	5	5.208,405	1.300,993	450,843	157,969	60,941	23,591	17,258

Selanjutnya hasil perhitungan statistik uji *chi-square*-nya yang dihitung menggunakan Persamaan 2.1 adalah sebagai berikut

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left\{\frac{(299.542 - 291.864{,}735)^2}{291.864{,}735} + \frac{(66.909 - 72.904{,}068)^2}{72.904{,}068} + \cdots + \frac{(0 - 17{,}258)^2}{17{,}258}\right\} = 5.366{,}929$$

Dari hasil perhitungan diatas maka diperoleh hasil statistik uji *chi-square* sebesar 5.366,929. Langkah selanjutnya adalah menentukan nilai *chi-square* tabel untuk mendapatkan kriteria uji pada pengujian ini. Dengan α = 0,05 dan derajat bebasnya adalah *(k*-1*)(p*+1*)* = (5-1)(5+1) = 24 maka nilai $\chi^2_{(1-0{,}05)}$adalah 36,4.

Hasil perhitungan uji *chi-square* ini menunjukan bahwa H_0 ditolak karena $\chi^2 > \chi^2_{(1-\alpha)}$atau 5.366,929 > 36,4. Artinya minimal ada dua kelompok populasi yang profil persentilnya tidak sama.

Uji Parsial

Setelah dilakukan uji keseluruhan atau uji secara simultan, hasil uji *chi-square* sebesar 5.366,929. Hal ini menunjukkan hipotesis nol di tolak, yang artinya minimal ada ada dua kelompok populasi yang profil persentilnya tidak sama. Kemudian untuk mengetahui populasi manakah yang distribusinya tidak sama dilakukan uji parsial menggunakan uji *chi-square.*

Kelompok ke-1 dengan Kelompok ke-2

Uji parsial ini bertujuan untuk membandingkan kelompok ke-1 dan kelompok ke-2. Hipotesis untuk pengujian tersebut adalah:

H_0 : $Q_1 = Q_2$ kelompok ke-1 dan kelompok ke-2 memiliki distribusi yang sama.

H_1 : $Q_1 \neq Q_2$ kelompok ke-1 dan kelompok ke-2 memiliki distribusi yang berbeda.

Melakukan pengujian kesamaan distribusi profil persentil pada uji parsial, dengan statistik uji *chi-square* sebagaimana menggunakan Persamaan 2.1 dan Persamaan 2.2. Untuk data kelompok ke-1 dan kelompok ke-2 tersaji pada Tabel 3.9.

Tabel 3.9 Data untuk Kelompok ke-1 dan Kelompok ke-2

Kelompok Populasi ke-	Bin ke-							Total
	0	1	2	3	4	5	6	
1	299.542	66.909	23.613	8.607	3.453	1.401	1.064	404.589
2	180.872	55.727	19.264	6.506	2.406	896	611	266.282
Total	480.414	122.636	42.877	15.113	5.859	2.297	1.675	670.871

Kemudian terlebih dahulu menentukan ekspektasi frekuensi setiap observasi dan dihitung menggunakan Persamaan 2.2. Berikut hasil perhitungan dari Persamaan 2.2.

$$E_{10} = \frac{T_{.0} \times T_{1.}}{T_{..}} = \frac{480.414 \times 404.589}{670.871} = 289.728{,}159$$

… dst

Hasil perhitungan ekspektasi frekuensi pada setiap observasi kelompok ke-1 dan kelompok ke-2 disajikan pada Tabel 3.10

Tabel 3.10 Hasil Perhitungan Nilai Ekspektasi untuk Kelompok ke-1 dan Kelompok ke-2

Kelompok Populasi ke-		Bin ke						
		0	1	2	3	4	5	6
1	O_{ij}	299.542	66.909	23.613	8.607	3.453	1.401	1.064
	E_{ij}	289.728,159	73.959,340	25.858,269	9.114,351	3.533,447	1.385,275	1.010,159
2	O_{ij}	180.872	55.727	19.264	6.506	2.406	896	611
	E_{ij}	190.685,841	48.676,660	17.018,731	5.998,649	2.325,553	911,725	664,841

Selanjutnya hasil perhitungan statistik uji *chi-square*-nya yang dihitung menggunakan Persamaan 2.1 adalah sebagai berikut:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left\{\frac{(299.542 - 289.728,159)^2}{289.728,159} + \frac{(66.909 - 73.959,340)^2}{73.959,340} + \cdots + \frac{(611 - 664,841)^2}{664,841}\right\} = 3.105,381$$

Dari hasil perhitungan diatas maka diperoleh hasil statistik uji *chi-square* sebesar 3.105,381. Langkah selanjutnya adalah menentukan nilai *chi-square* tabel untuk mendapatkan kriteria uji pada pengujian ini. Dengan α = 0,05 dan derajat bebasnya adalah *(k*-1*)(p*+1*)* = (2-1)(5+1) = 6 maka nilai $\chi^2_{(1-0,05)}$adalah 12,6.

Hasil perhitungan uji *chi-square* ini menunjukan bahwa H_0 ditolak karena $\chi^2 > \chi^2_{(1-\alpha)}$atau 3.105,381> 12,6. Artinya kelompok ke-1 dan kelompok ke-2 memiliki distribusi yang berbeda.

Untuk membandingkan setiap kelompok lainnya memiliki distribusi yang sama atau tidak, dengan cara yang sama hasil perhitungan uji parsial untuk keseluruhan populasi diringkas dan tersaji pada Tabel 3.11.

Hasil perhitungan uji *chi-square* pada Tabel 3.11 menunjukan bahwa semua H_0 ditolak karena $\chi^2 > \chi^2_{(1-\alpha)}$dengan α = 0,05 dan nilai $\chi^2_{(1-\alpha)} = 12{,}6$. Artinya semua populasi memiliki distribusi yang berbeda satu sama lain. Jika setiap populasi memiliki distribusi yang sama maka data frekuensi klaim kendaraan bermotor di Indonesia digabungkan dan membentuk satu model. Namun jika setiap populasi memiliki distribusi yang berbeda maka pemodelan dilakukan pada masing-masing populasi.

Tabel 3.11 Hasil Perhitungan Uji Parsial Menggunakan Uji *Chi-square* dengan α = 0,05 dan *df* = 6

No	**Kelompok Populasi ke-**	**Nilai *Chi-Square***	**Kriteria Uji dengan** $\chi^2_{(1-0,05)} = 12,6$
1	1 dengan 2	3.105,381	Tolak H_0
2	1 dengan 3	232,308	Tolak H_0
3	1 dengan 4	138,220	Tolak H_0
4	1 dengan 5	1.296,336	Tolak H_0
5	2 dengan 3	1.087,738	Tolak H_0
6	2 dengan 4	482,006	Tolak H_0
7	2 dengan 5	1997,151	Tolak H_0
8	3 dengan 4	69,528	Tolak H_0
9	3 dengan 5	1.028,726	Tolak H_0
10	4 dengan 5	514,206	Tolak H_0

D. Kesimpulan

Dalam makalah ini telah dibahas prosedur untuk menguji kesamaan beberapa kelompok populasimenggunakan uji *chi-square.* Dimana datayang digunakan adalah data asuransi kendaraan bermotor kelompok 1, 2, 3,4, dan 5 di Indonesia. Berdasarkan hasil dan pembahasan dapat disimpulkan bahwaada beberapa kelompok data frekuensi klaim asuransi kendaraan bermotor memiliki profil persentilnya tidak sama atau distribusinya berbeda. Untuk mengetahui kelompok mana yang berbeda maka dilakukan uji parsial menggunakan *chi-square*. Ternyata setiap kelompok frekuensi klaim asuransi kendaraan bermotor memiliki distribusi yang berbeda.Oleh karena itu jika akan dilakukan pemodelan pada kelompok tersebut pemodelan dilakukan pada masing-masing populasi.

Daftar Pustaka

Bedrick, E.J. dan Hossain, A. (2013). *Conditional Tests for Homogeneity of Zero-Inflated Poisson and Poisson-Hurdle Distributions.* Computational Statistics and Data Analysis, 61, 99-106.

Boucher, J.P., Denuit, M., dan Guill'en, M. (2006). *Risk Classification for Claim Counts: A comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models*. North American Actuarial Journal, 11-14:110-13.

Conover, W.J. (1980). *Pratical Nonparametric Statistics.* Edisi Kedua. Willey-Interscirnce, United States of America.

Daniel, W.W. (1989). *Statistik Nonparametrik Terapan*. Diterjemahkan oleh: Alex Tri K.W. Jakarta : PT. Gramedia.

Hallstrom, A.P. (2010). *A Modified Wilcoxon Test for Non-Negative Distributions with a Clump of Zeros.* Statistics in Medicine, 29, 391-400.

Johnson, W.D., Beyl, R.A., Burton, J.H., dan Romer, J.E. (2015). *A Simple Chi-Square Statistic for Testing Homogeneity of Zero-Inflated Distributions*. Open Journal of Statistics, 5, 483-493.

Kitab Undang – Undang Hukum Dagang (KUHD), Bab IX Tentang Asuransi atau Pertanggungan pada Umumnya, Pasal 246.

Kementerian Keuangan Republik Indonesia Badan Pengawasan Pasar Modal dan

Lembaga Keuangan (2011). Peraturan Ketua Badan Pengawasan Pasar Modal dan Lembaga Keuangan Nomor: PER-04/BL/2011.

Klugman, S.A., Panjer, H.H., dan Wilmot, G. (2004). *Loss Models. From data to decisions.* Edisi Kedua.Willey-Interscirnce, New York.

Lachenbruch, P.A. (1976) *Analysis of Data with Clumping at Zero*. Biometrische Zeitschrift, 18, 351-356.

Republik Indonesia (1992), "Undang – Undang RI No. 2 Tahun 1992 Tentang Usaha Perasuransian. Jakarta.

Rizanti R. (2013). *Pemodelan Regresi Zero-Inflated Poisson pada Data Klaim Asuransi Kendaraan Bermotor di Indonesia.* Skripsi S1 Program Studi Statistika. Universitas Islam Bandung.

Tse, S.K., Chow, S.C., Lu, Q.S. dan Cosmatos, D. (2009). *Testing Homogeneity of Zero-Inflated Poisson Populations*. Biometrical Journal, 51, 159-170.

Walpole, R.E. (1992). *Pengantar Statistika.* Edisi ketiga, Jakarta : PT. Gramedia Pustaka Utama.

Yuen, H.K., Chow, S.C. dan Tse, S.K. (2015). *On Statistical Tests for Homogeneity of Two Bivariate Zero-Inflated Poisson Populations*. Journal of Biopharmaceutical Statistics, 25, 44-53.