

BAB III

METODE CHAID *EXHAUSTIVE*

3.1 CHAID *Exhaustive*

Metode CHAID *Exhaustive* dikemukakan oleh D. Biggs *et al.* (1991) yang merupakan evaluasi dari metode sebelumnya yaitu CHAID (Kass, 1980) untuk penyesuaian bonferroni pada jumlah kategori lebih besar dari empat kategori dari peubah penjelas.

Metode CHAID *Exhaustive* mengemukakan prosedur penyekatan dengan cara melihat seluruh kemungkinan penggabungan dari pasangan kategori, secara bertahap hingga tersisa dua kategori. Prosedur ini memberikan perhitungan nilai pengali *bonferroni* yang lebih kecil, sehingga memberikan tipe kesalahan I yang tak terlampaui kecil dan berakibat jenis kesalahan II tak terlampaui besar. Oleh karena itu, hal ini menjamin bahwa kategori yang tersisa merupakan gabungan kategori yang paling nyata (Biggs, 1991).

3.2 Variabel-Variabel dalam Metode CHAID *Exhaustive*

Variabel yang digunakan dalam metode CHAID *Exhaustive* terdiri atas variabel dependen dan variabel independen. Metode klasifikasi dalam CHAID *Exhaustive* dilakukan berdasarkan pada hubungan yang ada antara kedua variabel tersebut. Gallagher (2002) menunjukkan bahwa CHAID akan membedakan variabel-variabel independennya menjadi tiga bentuk berbeda, yaitu:

1. Monotonik

Kategori-kategori pada variabel monotonik dapat digabungkan, hanya jika keduanya berdekatan satu sama lain, yaitu variabel-variabel yang kategorinya mengikuti urutan aslinya (data ordinal). Contohnya adalah: usia dan pendapatan.

2. Bebas

Kategori-kategori pada variabel bebas dapat dikombinasikan atau digabungkan walaupun keduanya berdekatan atau tidak satu sama lain (data nominal). Contohnya adalah: pekerjaan, kelompok, etnik dan area geografis.

3. Mengambang (*floating*)

Kategori-kategori pada variabel mengambang dapat diperlakukan sama seperti layaknya variabel monotonik kecuali untuk kategori terakhir yaitu *missing value*, yang dapat dikombinasikan atau digabungkan dengan kategori manapun.

3.3 Algoritma CHAID *Exhaustive*

Secara garis besar algoritma CHAID *Exhaustive* mempunyai tiga tahapan yaitu tahap penggabungan (*Merging*), tahap pemisahan (*Spliting*) dan tahap penghentian (*Stoping*). Diagram pohon pada algoritma CHAID *Exhaustive* dimulai dari simpul akar (*root simpul*) melalui tiga tahapan penggabungan, pemisahan dan penghentian pada setiap simpul yang terbentuk dan secara berulang.

3.3.1 Tahap Penggabungan (*Merging*)

Tahap pertama dalam algoritma CHAID *Exhaustive* adalah tahap penggabungan. Pada tahap ini hal yang dilakukan yaitu memeriksa signifikansi dari masing-masing kategori pada variabel independen. Kemudian kategori-kategori yang tidak signifikan akan digabungkan. Tahap penggabungan untuk setiap kategori-kategori yang tidak signifikan adalah sebagai berikut:

- 1) Membentuk tabel kontingensi dua arah untuk masing-kategori.
- 2) Menghitung statistik *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabungkan menjadi satu. Prosedur pengujian *chi-square* adalah sebagai berikut:
 - a. Menentukan hipotesis, sebagai berikut :

H_0 = Tidak terdapat hubungan antara variabel independen dengan variabel dependen.

H_1 = Terdapat hubungan antara variabel independen dengan variabel dependen

- b. Menentukan taraf signifikansi (α) yaitu = 0.05
 - c. Menentukan nilai $\chi^2_{0.05; (b-1)(k-1)}$ dengan menggunakan tabel *chi-square*.
 - d. Menentukan statistik uji, yaitu $\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
 - e. Menentukan daerah penolakan, yaitu $\chi^2 > \chi^2_{0.05; (b-1)(k-1)}$ atau *p-value* $< \alpha$
 - f. Mengambil kesimpulan:
 1. Apabila χ^2 termasuk ke dalam daerah penolakan, maka H_0 ditolak
 2. Apabila χ^2 tidak termasuk ke dalam daerah penolakan, maka H_0 diterima
- 3) Selanjutnya adalah menentukan nilai *p-value* berpasangan secara bersamaan, untuk masing-masing nilai *chi-square* yang berpasangan. Untuk pasangan kategori yang tidak signifikan, hal yang dilakukan adalah menggabungkan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang mempunyai nilai *chi-square* berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal dan kemudian lanjutkan ke langkah nomor selanjutnya.
 - 4) Memeriksa kembali kesignifikansian kategori baru setelah digabungkan dengan kategori lainnya. Jika masih ada pasangan yang belum signifikan, ulangi langkah nomor 3. Jika semua signifikan lanjutkan ke tahapan selanjutnya.

3.3.2 Tahap Pemisahan (*Splitting*)

Langkah pemisahan dipergunakan untuk memilih kategori mana yang akan digunakan untuk *split node* (pemisahan titik) terbaik. Pemisahan dilakukan

dengan cara membandingkan nilai *p-value* yang diperoleh dari tahap penggabungan pada setiap kategori. Tahap pemisahan adalah sebagai berikut:

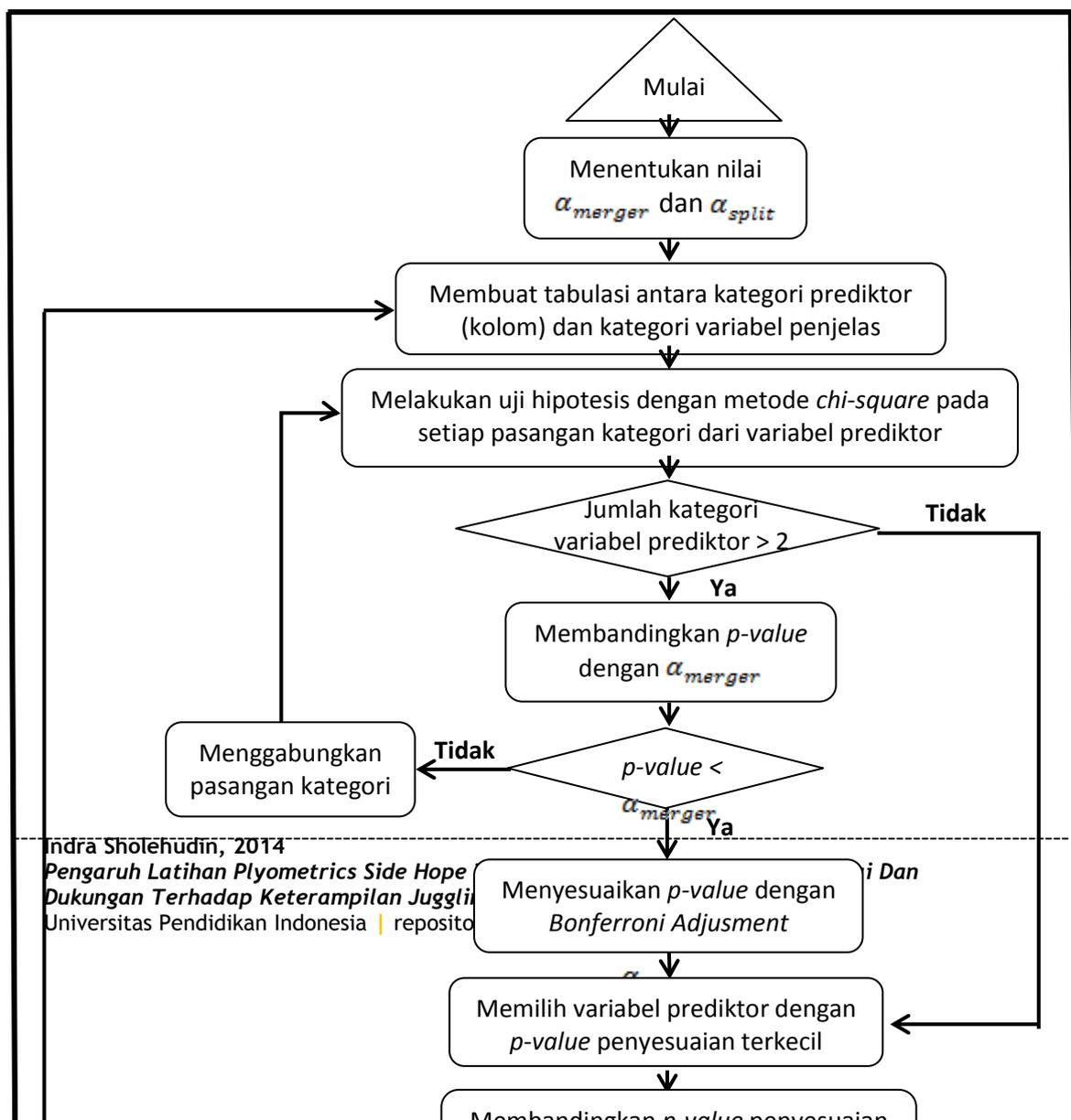
- 1) Memilih variabel independen yang memiliki nilai *p-value* terkecil, yang akan digunakan sebagai *split node*.
- 2) Jika nilai *p-value* \leq *alpha*, maka variabel independen dilakukan pemisahan. Jika tidak ada variabel independen yang mempunyai nilai *p-value* signifikan maka lanjutkan ke tahap selanjutnya dan variabel independen tersebut menjadi simpul akhir.

3.3.3 Tahap Pengehentian (*Stopping*)

Tahap penghentian dilakukan jika suatu proses pertumbuhan pohon harus dihentikan sesuai dengan peraturan penghentian di bawah ini:

- 1) Tidak ada variabel independen yang signifikan menunjukkan perbedaan terhadap variabel dependen.
- 2) Jika pohon telah mencapai batas nilai maksimum, maka proses pertumbuhan pohon akan dihentikan. Misalkan telah ditetapkan bahwa batas ke dalaman pertumbuhan pohon klasifikasi adalah 3, maka ketika pertumbuhan pohon sudah mencapai ke dalaman 3, pertumbuhan pohon klasifikasi dihentikan.

Untuk lebih jelasnya mengenai langkah-langkah dari perhitungan pada algoritma CHAID *Exhaustive*, dapat dilihat pada gambar 3.1 gambar alur perhitungan CHAID *Exhaustive*.



Gambar 3.1. Algoritma CHAID *Exhaustive*

3.4 Uji *Chi-Square*

Uji *Chi-Square* digunakan untuk menentukan apakah ada perbedaan yang signifikan antara frekuensi yang diharapkan dengan frekuensi hasil observasi dari satu atau lebih kategori. Tabel kontingensi adalah analisis data untuk melihat hubungan antara beberapa variabel dalam satu tabel. Variabel yang dianalisis merupakan variabel kategorikal, yang memiliki skala nominal atau ordinal. Penentuan profil segmen/*cluster* diperoleh dari hasil tabel kontingensi dua arah. Secara umum tabel disajikan sebagai berikut:

Tabel 3.1 Tabel Kontingensi Dua Arah

Kolom baris	1	2	...	c	Total
------------------------	----------	----------	------------	----------	--------------

1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
·	·	·	...	·	·
·	·	·	...	·	·
·	·	·	...	·	·
R	n_{r1}	n_{r2}	...	n_{rc}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.c}$	N

Statistik uji yang digunakan adalah chi-Square sebagai berikut (Agresti, 2002)

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

dengan derajat bebas : $(r-1)(c-1)$

O_{ij} : observasi baris ke-i dan kolom ke-j

e_{ij} : frekuensi harapan baris ke-i dan kolom ke-j

r : banyaknya baris

c : banyaknya kolom

Untuk menghitung frekuensi harapan masing-masing sel digunakan rumus sebagai berikut:

$$e_{ij} = \frac{r_i c_j}{n}$$

di mana, r_i : jumlah baris ke-i

c_j : jumlah kolom ke-j

n : jumlah pengamatan

Pengujian dilakukan dengan membandingkan nilai χ^2 hitung dengan nilai χ^2 tabel dengan derajat bebas adalah $(r-1)(c-1)$ untuk tabel ukuran $(r \times c)$. Jika nilai χ^2 hitung lebih besar dari nilai χ^2 tabel, maka diputuskan hipotesis awal ditolak. Nilai koefisien kontingensi yang diperoleh berkisar antara 0 dan 1. Nilai perhitungan uji *Chi-Square* harus memenuhi syarat maksimal 20% dari jumlah sel yang ada memiliki nilai ekspektasi harapan tiap sel kurang dari 5, data yang tidak memenuhi asumsi di atas digunakan uji *fisher's exact*.

Tabel 3.2 Probabilitas Kejadian

Kolom baris	1	2	...	c	Total
1	p_{11}	p_{12}	...	p_{1c}	$p_{1.}$
2	p_{21}	p_{22}	...	p_{2c}	$p_{2.}$
.
.
.
R	p_{r1}	p_{r2}	...	p_{rc}	$p_{r.}$
Total	$p_{.1}$	$p_{.2}$...	$p_{.c}$	P

Dalam pengambilan keputusan H_0 ditolak jika χ^2 masuk ke dalam daerah penolakan. Sebaliknya H_0 diterima jika χ^2 tidak masuk ke dalam daerah penolakan.

3.5 Uji Hipotesis

Uji hipotesis merupakan bagian yang sangat penting dalam statistika inferensi. Hipotesis yang bersifat statistik sebetulnya dapat diartikan sebagai suatu asumsi mengenai parameter fungsi frekuensi variabel random. Uji hipotesis merupakan suatu prosedur untuk menentukan apakah suatu hipotesis diterima atau tidak (Levin & Rubin, 1998).

Hipotesis yang digunakan untuk melakukan pengujian adalah hipotesis nol (H_0) dan hipotesis alternatif (H_1). Hipotesis nol merupakan hipotesis yang akan diuji dan nantinya akan diterima atau ditolak tergantung pada hasil eksperimen. Hipotesis nol menampilkan suatu teori yang dipercaya kebenarannya atau karena digunakan sebagai dasar argumen, namun belum terbukti. Hipotesis alternatif merupakan pernyataan berupa apa yang uji hipotesis statistik munculkan di mana hipotesis alternatif akan diterima jika hasil eksperimen menolak H_0 .

Pada algoritma CHAID *Exhaustive*, uji hipotesis dilakukan pada saat penggabungan variabel independen dan pembagian simpul. Hipotesis nol pada penggabungan variabel independen adalah pasangan variabel independen tidak memiliki perbedaan dalam memberi pengaruh pada variabel target, sehingga dapat digabungkan menjadi susunan tunggal. Sementara yang menjadi hipotesis alternatif adalah keadaan di mana hipotesis nol ditolak, yaitu pasangan variabel memiliki perbedaan dalam memberikan pengaruh pada variabel target, sehingga tidak dapat digabungkan. Pada saat pembagian simpul yang menjadi hipotesis nol adalah kategori-kategori yang ada di dalamnya bersifat homogen, sehingga simpul tidak perlu dibagi. Sementara hipotesis alternatifnya berupa keadaan yang menolak hipotesis nol, yaitu kategori-kategori yang terdapat pada simpul tersebut bersifat heterogen, sehingga terjadi pembagian simpul.

Hasil uji hipotesis ini dinyatakan dalam tingkat signifikansi. Signifikansi merupakan istilah statistik untuk menyatakan seberapa yakin hubungan atau perbedaan yang ada antar variabel. Apabila uji hipotesis menunjukkan hasil menerima hipotesis nol, maka tidak terdapat perbedaan yang signifikan antara data yang diuji terhadap hipotesis yang ada. Jika uji hipotesis menerima hipotesis alternatif, maka terdapat perbedaan yang signifikan antara data yang diuji dengan hipotesis, sehingga hipotesis alternatif diambil sebagai kesimpulan dari hasil uji hipotesis.

Dalam uji hipotesis terdapat dua jenis kesalahan yang mungkin terjadi. Kesalahan tipe I (α) adalah kesalahan yang terjadi ketika menolak hipotesis nol, di mana hipotesis tersebut benar. Sedangkan kesalahan tipe II (β) adalah kesalahan yang terjadi ketika menerima hipotesis nol, di mana hipotesis tersebut salah.

Tabel 3.3 Uji Hipotesis

		Keputusan	
		Menolak H_0	Menerima H_0
Kebenaran	H_0	Kesalahan tipe I	Keputusan yang benar
	H_1	Keputusan yang benar	Kesalahan tipe II

Terdapat dua jenis uji pada uji hipotesis, yaitu uji satu arah dan uji dua arah. Pada uji satu arah, nilai-nilai yang ada ketika menolak hipotesis nol berada seluruhnya di satu arah, nilai-nilai yang ada ketika menolak hipotesis nol berada seluruhnya di satu arah distribusi probabilitas. Sementara uji dua arah, nilai-nilai yang ada ketika menolak hipotesis nol berada seluruhnya di kedua arah distribusi probabilitas. Hal ini dapat dilihat melalui gambaran berikut ini:

UJI SATU ARAH

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0 \text{ (atau } H_1: \mu < \mu_0 \text{)}$$

UJI DUA ARAH

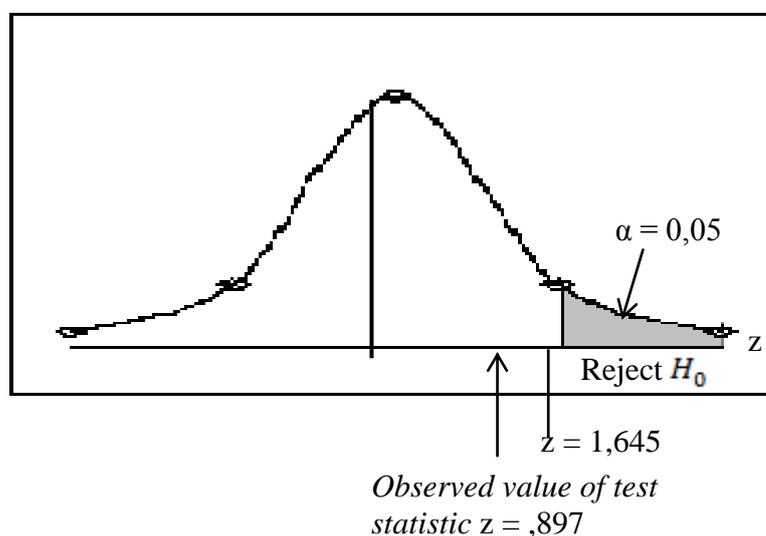
$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

3.6 Tingkat Signifikan dan *P-value*

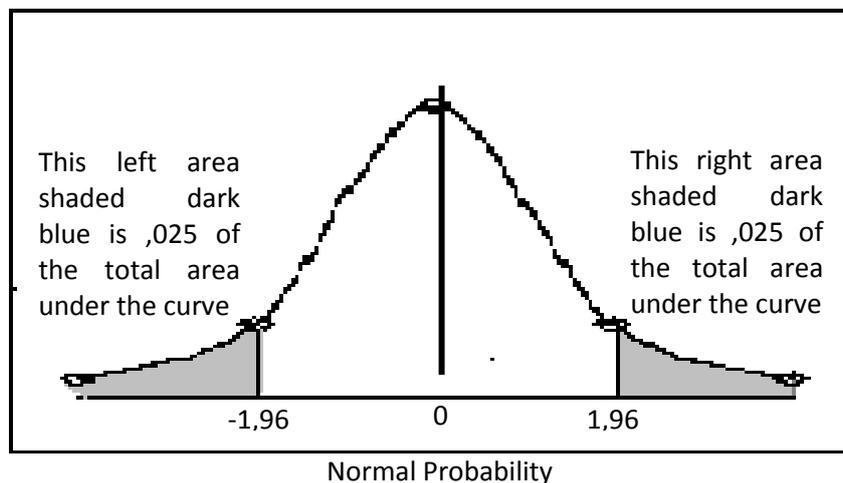
Tingkat signifikan (α) merupakan probabilitas tertentu menolak hipotesis nol ketika hipotesis tersebut benar. Tingkat signifikan ini adalah kesalahan tipe I yang ditentukan oleh investigator yang berhubungan dengan konsekuensi terjadinya kesalahan. Tingkat signifikan yang paling sering digunakan adalah 0,05 (5%).

Nilai probabilitas (*p-value*) dari uji hipotesis statistik adalah besarnya peluang melakukan kesalahan apabila kita memutuskan untuk menolak hipotesis nol (Kurniawan, 2008).



Gambar 3.2. Uji Hipotesis Satu Arah pada $\alpha = 0,05$

Dalam hal ini apabila $p\text{-value} < \alpha$, maka uji hipotesis menunjukkan menolak hipotesis nol, dan sebaliknya apabila $p\text{-value} > \alpha$, maka uji hipotesis menunjukkan menerima hipotesis nol. Hal ini menunjukkan, $p\text{-value}$ yang semakin mendekati nol, maka akan cenderung menolak hipotesis nol. Demikian pula sebaliknya, $p\text{-value}$ yang semakin mendekati angka satu, maka cenderung menerima hipotesis nol.



Gambar 3.3 Uji Hipotesis Dua Arah pada $\alpha = 0,05$

3.7 Koreksi *Bonferroni*

Koreksi *Bonferroni* merupakan penyesuaian statistik untuk perbandingan berganda. Dalam menguji n kejadian atau perbandingan dari n variabel, nilai tingkat signifikan akan dibagi dengan n . Pengujian akan dilakukan pada nilai α yang sudah dibagi tersebut. Dengan demikian akan menjamin, bahwa keseluruhan kesepakatan untuk terjadinya kesalahan tipe I masih kurang dari nilai α .

Koreksi *Bonferroni* juga dapat diterapkan dengan menyesuaikan *p-value*. Hal ini dilakukan dengan mengalikan sejumlah kejadian yang akan diuji. Jika *p-value* yang telah disesuaikan ini lebih besar dari 1,0 maka akan dibulatkan ke bawah menjadi 1.

Koreksi *Bonferroni* pada algoritma CHAID *Exhaustive* digunakan untuk penyesuaian *p-value* dari variabel prediktor yang telah terjadi pengurangan kategori. Penyesuaian ini dilakukan dengan melakukan pengalihan antara *p-value*

sebelum terjadi pengurangan kategori dengan pengali *Bonferroni*. Jika jumlah kategori semula adalah c , dan jumlah kategori setelah terjadi penggabungan adalah r , maka nilai pengali *bonferroni* yang diberikan oleh CHAID *Exhaustive* adalah sebagai berikut:

Dalam metode ini penggabungan dilakukan bertahap sejak langkah ke $k = 1$ hingga tersisa dua kategori pada $k = c-2$, maka

$$N_{BC}(c) = 1 + \sum_{k=1}^{c-2} N(c-k+1, c-k)$$

di mana untuk:

1. Variabel ordinal:

$$N(c-k+1, c-k) = c-k$$

2. Variabel nominal:

$$N(c-k+1, c-k) = \sum_{i=0}^{(c-k)-1} (-1)^i \frac{((c-k)-i)^{c-k+1}}{i!((c-k)-i)!}$$

3. Variabel mengambang:

Jika kategori mengambang tidak dapat digabungkan maka koefisien *Bonferroni* sama dengan penyesuaian *Bonferroni* pada skala ordinal:

$$N_{BC}(c-1) = 1 + \sum_{k=1}^{c-3} N(c-k+1, c-k)$$

Jika kategori mengambang dapat digabungkan maka koefisien *Bonferroni* untuk tipe ini sama dengan koefisien *Bonferroni* tanpa kategori mengambang ($c-1$ kategori) untuk skala ordinal ditambah jumlah gabungan bila kategori mengambang disertakan.

$$N_{BC} = N_{BC}(c-1) + (c-1) + 1$$