MODUL 5

Analisis Data Kategorik

Prof. Dr. Zanzawi Soejoeti

PENDAHULUAN

Sering kali informasi sampel yang dikumpulkan oleh peneliti mempunyai bentuk: *data kategorik.* Istilah ini berarti, observasi hanya diklasifikasi ke dalam kategori-kategori sehingga himpunan data berbentuk cacah frekuensi untuk kategori-kategori itu. Data semacam itu banyak sekali terjadi dalam hampir semua bidang studi kuantitatif, khususnya dalam ilmu-ilmu sosial. Dalam suatu studi tentang agama yang dianut sekumpulan orang akan diklasifikasi ke dalam agama Islam, Protestan, Katolik, Hindu, dan sebagainya. Dalam suatu survei tentang kecocokan pekerjaan, orang yang bekerja dapat diklasifikasi sebagai puas, netral, dan tidak puas dengan pekerjaan mereka. Dalam perkembangbiakan tanaman, turunan dari persilangan tanaman dapat dikelompokkan ke dalam beberapa genotif. Benda hasil produksi dapat dibagi-bagi dalam beberapa kategori, seperti *bebas cacat, sedikit cacat* dan *ditolak*. Dalam semua contoh-contoh ini tiap kategori didefinisikan sebagai sifat kualitatif. Kategori dapat juga didefinisikan dengan mengatakan rentang nilai sebarang pada skala pengukuran numerik yang asli, seperti penghasilan yang dikategorikan sebagai tinggi, sedang atau rendah; curah hujan yang diklasifikasikan sebagai berat, sedang atau ringan. Mengembangkan pemikiran ini lebih lanjut, distribusi frekuensi pengukuran yang dikelompokkan dalam interval kelas dapat juga divisualisasi sebagai bentuk data kategorik yang interval kelasnya mendefinisikan kategori-kategori itu meskipun bentuk data terakhir ini secara tradisional diberi nama *data numerik* atau *data pengukuran.* Prosedur inferensi tertentu yang dirancang untuk data kategorik masih dapat diterapkan untuk itu.

KEGIATAN BELAJAR 1

Uji Kesesuaian

Bentuk data kategorik yang paling sederhana hanya memuat dua kategori yang didefinisikan sebagai mempunyai atau tidak mempunyai sifat tertentu. Peranan yang dimainkan distribusi binomial dalam memodelkan data dikotomi ini pernah kita pelajari dalam *Metode Statistik 1*, demikian juga prosedur inferensi yang berkaitan dengan itu. Tujuan kita di sini adalah menyajikan beberapa prosedur yang dapat digunakan untuk mempelajari data yang dikelompokkan dalam beberapa kategori. Beberapa contoh akan membantu melukiskan sifat inferensinya dan untuk menyiapkan analisis kita yang akan datang.

Contoh 5.1

Pilihan konsumen: Mesin cuci merek A dijual dalam lima warna yang berbeda, dan peneliti pasar ingin mempelajari popularitas berbagai warna itu. Frekuensi yang disajikan dalam Tabel 5.1 diamati dari sampel random 300 penjualan baru-baru ini. Sebagai langkah awal, peneliti mungkin ingin menguji hipotesis bahwa kelima warna itu popularitasnya sama.

Tabel 5.1
Data Penjualan 5 Warna Mesin Cuci Merek A

Warna Avokat	Coklat Muda	Merah	Biru	Putih	Jumlah
88	65	52	40	55	300

Contoh 5.2

Model genetika: Turunan yang dihasilkan oleh persilangan antara dua jenis tanaman tertentu dapat dengan A, B, dan C. Model teoretis keturunan menegaskan bahwa keturunan jenis A, B, dan C harus dalam perbandingan 1 : 2 : 1. Untuk verifikasi percobaan 90 tanaman dikembangkan oleh persilangan dua jenis tanaman tertentu itu. Klasifikasi genetika tanaman-tanaman itu dicatat dalam Tabel 5.2. Apakah data itu mendukung atau kontradiksi dengan model genetika itu?

Tabel 5.2
Klasifikasi Tanaman Biak Silang

Genotif			Jumlah
A	B	C	
18	44	28	90

Contoh 5.3

Distribusi Poisson untuk klaim asuransi: Seorang ahli asuransi ingin mempelajari frekuensi klaim asuransi untuk perawatan medis dalam rumah sakit di antara keluarga dengan dua anak yang ayah dan ibu keduanya berumur di bawah 50 tahun. Ahli itu mengambil sampel dengan 200 keluarga seperti itu dan mencatat distribusi frekuensi banyak klaim ini selama periode empat tahun. Dari himpunan data yang tertuang dalam Tabel 5.3, apakah distribusi Poisson tampak menggambarkan data ini dengan baik?

Tabel 5.3
Distribusi Frekuensi Banyak Klaim untuk Perawatan Medis dalam Rumah Sakit

Banyak klaim	0	1	2	3	4	5	6	7	Jumlah
Frekuensi	22	53	58	39	20	5	2	1	200

Contoh 5.4

Distribusi normal untuk skor penampilan: Data berikut adalah sampel random berukuran 20 tentang skor penampilan dalam lomba tertentu. Data telah diurutkan dari kecil ke besar

Tabel 5.4
Skor Penampilan yang Diurutkan

16,7	17,4	18,1	18,2	18,8	19,3	22,4	22,5	24,0	24,7
25,9	27,0	35,1	35,8	36,5	37,6	39,8	42,1	43,2	46,2

Kita ingin menguji bahwa sampel random ini merupakan observasi pada variabel random berdistribusi normal dengan mean 30 dan variansi 100.

A. UJI KESESUAIAN DISTRIBUSI MULTINOMIAL

Untuk membentuk model peluang yang dapat menampung jenis data dalam beberapa contoh sebelumnya, kita harus memperluas model *Bernoulli trials* sehingga hasilnya dapat diklasifikasi ke dalam beberapa kategori karena trials semacam itu dinamakan *trials multinomial.*

Trial multinomial:
1. Hasil tiap trial masuk ke dalam salah satu dari k kategori atau sel yang saling pisah, ditulis 1, 2, …, k
2. Peluang bahwa trial akan menghasilkan dalam sel ke-i ditulis dengan $p_i\,; i = 1, ..., k,$ dan $\sum_{i=1}^{k} p_i = 1$. Peluang sel ini tetap sama untuk semua trial.
3. Trial-trial itu independen.

Untuk kasus khusus k = 2, definisi trial multinomial yang formal ini serupa dengan definisi Bernoulli trials. Syarat-syarat yang mendefinisikan trial multinomial dipenuhi dengan baik dalam situasi percobaan yang trial-trialnya diulang-ulang secara independen di bawah kondisi yang sama. Jika sampel random dari suatu populasi yang terdiri dari elemen-elemen dalam beberapa kategori, syarat-syarat itu dipenuhi jika pengambilan sampel dilakukan dengan pengembalian dan akan *mendekati dipenuhi* jika hanya bagian kecil dari populasinya yang diambil sebagai sampel meskipun pengambilan sampel dilakukan tanpa pengembalian.

Untuk sederetan n pengulangan suatu trial multinomial, frekuensi untuk sel 1, 2, …, k ditulis dengan n_1, n_2, …, n_k. Sebelum trial itu benar-benar dilakukan, ini adalah variabel random yang dapat mengambil nilai bilangan bulat tidak negatif, dan jumlahnya harus memenuhi $n_1 + n_2 + \ldots + n_k = n$. Struktur dasar ini disajikan dalam Tabel 5.5.

Tabel 5.5
Struktur Data Multinomial

Sel	1	2	...	k	Jumlah
Peluang	p_1	p_2	...	p_k	1
Frekuensi dalam n trials	n_1	n_2	...	n_k	n

Distribusi Multinomial

Distribusi peluang bersama frekuensi-frekuensi sel $n_1,...,n_k$, dalam *n* trial multinomial dinamakan *distribusi multinomial dengan parameter* $\rho_1,...,\rho_k$, yang masing-masing merupakan peluang sel. Fungsi peluangnya adalah:

$$f(n_1,...,n_k) = \frac{n!}{n_1!\,...\,n_k!}\,p_1^{n_1}\,...\,p_k^{n_k} \text{ untuk } \sum_{i=1}^{k} n_i = n$$

Parameter-parameter itu memenuhi $\sum_{i=1}^{k} p_i = 1$

Rumus matematik untuk fungsi peluang multinomial tidak terlalu penting dalam pembicaraan kita di sini. Tetapi beberapa sifat distribusi itu pantas mendapatkan perhatian kita di sini. Pertama, k variabel n_1, n_2, …, n_k dibatasi oleh $\sum_{i=1}^{k} n_i = n$ sehingga jika nilai-nilai (k-1) variabel yang mana pun telah tertentu maka nilai variabel sisanya otomatis tertentu pula. Sifat ini dapat dinyatakan dengan mengatakan bahwa derajat bebas yang berkaitan dengan n_1, n_2, …, n_k adalah (k-1). Pernyataan yang sama berlaku juga untuk himpunan parameter p_1, p_2, … p_k dengan pembatasan $\sum_{i=1}^{k} p_i = 1$. Kedua, kita menunjuk sel pertama "sukses" dan mengelompokkan sel sisanya ke dalam satu sel yang kita namakan "gagal" maka variabel random n_1 dapat dikatakan sebagai banyak sukses dalam n Bernoulli trials yang distribusinya binomial (n; p_1). Dengan demikian, $E(n_1) = np_1$, var $(n_1) = np_1(1 - p_1)$. Dengan alasan yang sama bagi setiap sel yang ada maka kita peroleh

$$E(n_i) = np_i \,;\; \mathrm{Var}(n_i) = np_i\,(1 - p_i)\,;\; i = 1, 2, ..., k$$

maka kita peroleh hasil pertama:

Frekuensi sel harapan = n (peluang sel)

yang akan memainkan peranan penting dalam pengembangan uji kesesuaian yang akan kita pelajari di bawah.

Oleh karena jumlah semua n_i tetap maka tidak mengherankan jika kita mendapatkan kovariansi yang negatif. Bentuknya adalah:

$$\text{kov}\,(n_i, n_j) = -\,np_ip_j\ ;\ i \neq j$$

Sekarang kita akan melakukan uji apakah model yang diberikan dalam hipotesis nol sesuai dengan data yang ada, dan ini dinamakan *uji kesesuaian.*

Peluang sel sepenuhnya ditentukan dalam H_0

Hipotesis nol yang menentukan sepenuhnya peluang sel berbentuk:

$H_0 : p_i = p_{10}, \ldots, p_k = p_{k0}$

Dengan $p_{10}, \ldots, p_{k0}$ adalah bilangan-bilangan diberikan yang memenuhi $p_{10} + \ldots + p_{k0} = 1$. Kembali ke contoh 5.1, hipotesis nol bagi peneliti pasar bahwa lima warna semuanya sama populernya dapat diterjemahkan menjadi $H_0 : p_1 =\ H_0 = \frac{1}{5}, \ldots, p_s = \frac{1}{5}$. Dalam contoh 5.2 hipotesis nol mempunyai bentuk $H_0 : p_1 = \frac{1}{4}; \quad p_2 = \frac{2}{4}; \ p_3 = \frac{1}{4}$.

Setelah peluang sel ditentukan, frekuensi sel harapan segera dapat dihitung dengan *mengalikan* peluang-peluang ini dengan ukuran sampel n. Uji kesesuaian berupaya untuk menentukan apakah ada perbedaan yang mencolok antara frekuensi sel pengamatan dan frekuensi harapan di bawah H_0. Ukuran yang berguna untuk perbedaan keseluruhan adalah:

$$x^2 = \sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{sel} \frac{(O-E)^2}{E}$$

dengan O dan E masing-masing melambangkan frekuensi pengamatan (observasi) dan frekuensi harapan yang bersesuaian. Perbedaan dalam tiap sel diukur dengan kuadrat selisih antara frekuensi pengamatan dan harapan dibagi oleh frekuensi harapan. Ukuran χ^2 adalah jumlah kuantitas ini untuk semua sel.

Statistik χ^2 mula-mula diusulkan oleh Karl Pearson (1857 – 1936) yang menemukan untuk n besar distribusi yang mendekati distribusi khi-kuadrat dengan derajat bebas (k-1). Karena distribusi ini, statistik itu ditulis dengan χ^2 dan dinamakan *Statistik* χ^2 *Pearson untuk uji kesesuaian.* Oleh karena nilai yang besar dari perbedaan keseluruhan menunjukkan ketidaksesuaian antara data dan hipotesis, ekor atas distribusi χ^2 merupakan daerah penolakan.

Uji kesesuaian χ^2 *Pearson*: Hipotesis nol : $H_0 : p_1 = p_{10}, \ldots, p_k = p_{k0}$ Statistik Penguji : $\chi^2 = \sum \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum \frac{(O-E)^2}{O}$ Distribusi : Jika n cukup besar sehingga tidak ada frekuensi sel harapan yang terlalu kecil, statistik pengujinya mendekati distribusi χ^2 dengan db = (k – 1) = banyak sel – 1. Daerah penolakan : $\chi^2 \geq \chi^2_{\alpha}$, titik α atas distribusi χ^2 dengan db = (k - 1)

Harus diingat bahwa uji kesesuaian χ^2 Pearson adalah uji pendekatan yang hanya berlaku untuk sampel-sampel besar. Sebagai aturan, n harus cukup besar sehingga frekuensi harapan dalam tiap sel paling sedikit 5. Beberapa statistisi menganjurkan bahwa frekuensi harapan dapat diturunkan menjadi kira-kira 1 pada sel akhir distribusi bermodus satu, seperti distribusi binomial, Poisson atau normal.

Contoh 5.5

Dari kata yang diberikan dalam Contoh 5.1, ujilah hipotesis nol bahwa semua 5 warna itu mempunyai kepopuleran yang sama, dengan tingkat signifikansi $\alpha = 0{,}05$.

Di sini hipotesis nol $H_0 : p_1 = p_2 = \ldots = p_5 = 0{,}2$. Hitungan-hitungan yang kita lakukan untuk statistik penguji tertuang dalam Tabel 5.6, dengan frekuensi harapan dihitung dengan mengalikan peluang sel dengan n = 300.

Tabel 5.6
Uji Kesesuaian untuk daxta Tabel 5.1

Sel	Warna Avokat	Coklat Muda	Merah	Biru	Putih	Jumlah
Frek Pengamatan (0)	88	65	52	40	55	300
Peluang di bawah H_0	0,2	0,2	0,2	0,2	0,2	1
Frek Harapan (E)	60	60	60	60	60	300
$\frac{(O-E)^2}{E}$	$\frac{28^2}{60}$	$\frac{5^2}{60}$	$\frac{8^2}{60}$	$\frac{(20)^2}{60}$	$\frac{5^2}{60}$	
	13,067	0,417	1,067	6,667	0,417	21,635 = χ^2 db = 4

Titik 5% atas distribusi χ^2 dengan db = 4 adalah 9,487. Oleh karena nilai χ^2 pengamatan (hitungan) jauh lebih besar dari χ^2 tabel maka hipotesis nol ditolak pada $\alpha =$ 0,05. Dengan demikian, ada perbedaan yang signifikan antara pengamatan dan hipotesis yang menyatakan bahwa kelima warna itu mempunyai popularitas yang sama. Tampak, fakta cukup kuat menolak H_0 dengan $\alpha = 0{,}05$.

Untuk memahami lebih jauh sifat penyimpangan dari model peluang sel sama, kita perhatikan kontribusi masing-masing sel dalam nilai χ^2 dalam baris terakhir Tabel 5.5. Kontribusi terbesar untuk statistik χ^2 dibuat oleh warna avokat, yang selisih (O – E) positif; kontribusi yang besar yang lain untuk statistik penguji ini dibuat oleh biru, yang penjualannya di bawah dari yang diharapkan.

Contoh 5.6

Dari data Contoh 5.2 kita uji hipotesis bahwa turunan persilangan tanaman itu mengikuti teori keturunan. Jadi $H_0 : p_1 = \frac{1}{4};\ p_2 = \frac{2}{4};\ p_3 = \frac{1}{4}$.

Kita hitung frekuensi harapan: $A : \frac{1}{4}(90) = 22,5;$ $B : \frac{2}{4}(90) = 45$ $C : \frac{1}{4}(90) = 22,5$. Sehingga

$$X^2 : \frac{(18-22,5)^2}{22,5} + \frac{(44-45)^2}{45} + \frac{(28-22,5)^2}{22,5} = 0,9 + 0,02 + 1,34$$

$$= 2,26$$

Titik 5% atas distribusi χ^2 dengan db = 2 adalah 5,9915.

Jadi, H_0 tidak ditolak.

B. UJI KESESUAIAN DISTRIBUSI YANG LAIN

Dalam menghitung frekuensi sel harapan sering kali memerlukan pengetahuan tentang nilai beberapa parameter yang tidak dinyatakan oleh hipotesis nol. Contoh 5.3 melukiskan keadaan seperti itu; di situ kita harus mengetahui nilai λ (yang merupakan mean distribusi Poisson) untuk menghitung peluang dari tabel Poisson. Demikian juga, dalam menguji apakah data merupakan sampel dari distribusi normal, nilai μ dan σ diperlukan untuk menghitung peluang. Dalam keadaan seperti itu, pertama-tama kita harus menaksir nilai parameter-parameter yang tidak diketahui itu dari data, dan menggunakan nilai taksiran ini untuk menghitung peluang sel. Selanjutnya, frekuensi harapan dan statistik penguji $\times^2$ dihitung dengan cara yang serupa, seperti yang kita pelajari di atas; tetapi derajat bebas untuk $\times^2$ berkurang lagi dengan banyak parameter yang ditaksir.

Sebelum kita menghitung statistik $\times^2$ kita harus memeriksa kembali apakah ada frekuensi sel harapan yang terlalu kecil. Oleh karena keabsahan uji kesesuaian kita menghindarkan keadaan ini. Sel yang berdekatan mungkin harus disatukan sampai frekuensi harapan di dalamnya paling sedikit 5.

Statistik $\times^2$ dihitung dari sel yang telah diubah itu dan banyak derajat bebas diberikan oleh

db = (banyak sel) – 1 – (banyak parameter yang ditaksir)

Contoh 5.7

Lakukan uji apakah model Poisson cukup baik menggambarkan data yang tertuang dalam Tabel 5.3.

Di sini hipotesis nol kesesuaian distribusi Poisson meninggalkan masalah parameter λ yang tidak dinyatakan nilainya. Kita ingat bahwa λ adalah mean distribusi Poisson. Dengan demikian, kita dapat menaksirnya dengan mean sampel yang kita hitung dari Tabel 5.3, yakni

$$\begin{aligned}\overline{x} &= \frac{\sum \text{(nilai x frekuensi)}}{n} \\ &= \frac{0(22) + 1(53) + ... + 7(1)}{200} \\ &= \frac{410}{200} = 2,05 \approx 2,0\end{aligned}$$

Dengan melihat pada tabel distribusi Poisson dengan $\lambda = 2$ kita peroleh peluang sel, yang jika kita kalikan dengan n = 200 akan kita peroleh frekuensi harapan. Nilai-nilai ini kita sajikan dalam Tabel 5.7 dengan penghitungan lengkap untuk statistik χ^2. Oleh karena jumlah dua fungsi harapan terakhir lebih kecil dari 5, sel-sel ini kita satukan dengan sel sebelumnya. Dengan mengambil $\alpha = 0{,}05$, titik 5% atas distribusi χ^2 dengan db = 4 adalah 9,488, yang lebih besar dari nilai χ^2 hitungan 2,33. Dengan demikian, hipotesis nol tidak ditolak pada $\alpha = 0{,}05$, dan dapat kita simpulkan bahwa model Poisson tidak kontradiksi dengan data itu.

Tabel 5.7
Uji kesesuaian χ^2 model Poisson data Tabel 5.3

Banyak Klaim	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Jumlah**
Frekuensi pengamatan (0)	22	53	58	39	20	5	2 8	1	200
Peluang Poisson dengan $\lambda = 2$	0,135	0,271	0,271	0,180	0,090	0,036	0,012	0,005	1,00
Frekuensi harapan (E)	27,0	54,2	54,2	36,0	18,0	7,2	2,4 10	1,0	200
$\frac{(O-E)^2}{E}$	0,926	0,027	0,266	0,250	0,222		0,638		2,33 = χ^2
									db = 6-1-1 = 4

Catatan:

Uji lain yang dapat kita lakukan di sini adalah uji yang didasarkan atas fakta bahwa variansi distribusi Poisson sama dengan *mean*-nya. Jika $s^2/\bar{x}$ jauh berbeda dengan 1, yakni < $\chi^2_{\ 0,975}/(n-1)$ atau > $\chi^2_{\ 0,025}/(n-1)$ hipotesis bahwa model Poisson sesuai dengan data ditolak. Daerah penolakan pendekatan ini didasarkan atas statistik χ^2 dengan db = (n-1).

Dalam setiap aplikasi uji kesesuaian dianjurkan bahwa parameter-parameter ditaksir dari data yang dikelompokkan dengan satu dari dua cara yang khusus, yaitu *metode kemungkinan terbesar* atau *metode khi-kuadrat terkecil.*

Contoh 5.8

Dari data contoh 5.4 kita ingin menguji bahwa data itu merupakan observasi pada variabel random berdistribusi normal dengan mean 30 dan variansi 100. Untuk ini kita sekarang memutuskan membentuk empat kelas (sel) dengan frekuensi sel harapan yang sama. Empat kelas ini dibentuk dengan menghitung kuartal pertama (K_1), kedua (K_2), dan ketiga (K_3). Dengan mengingat:

$$K = \mu + \sigma z_p$$

dengan Z_p adalah variabel random normal standar yang membatasi luasan 100% di sebelah kiri. Jadi $Z_{0,25}$ = -0,67; $Z_{0,50}$ = 0; dan $Z_{0,75}$ = 0,67 sehingga

$K_1 = 30 + 10(-0,67) = 23,3$

$K_2 = 30$

$K_3 = 30 + 10(0,67) = 36,7$

Sel 1 memuat observasi yang lebih kecil atau sama dengan 23,3; sel 2 memuat observasi antara 23,3 dan 30 (termasuk 30); dan sel 3 memuat observasi antara 30 dan 36,7 (termasuk 36,7) dan sisa observasi masuk dalam sel 4.

Maka, kita peroleh klasifikasi data sebagai berikut.

	Sel 1 **(-∞; 23,3]**	**Sel 2** **(23,3 ; 30]**	**Sel 3** **(30; 36,7]**	**Sel 4** **(36,7 ; ∞)**	**Jumlah**
Frekuensi pengamatan	8	4	3	5	20
Frekuensi harapan	5	5	5	5	20

Kita hitung statistik penguji χ^2, kita peroleh:

$$x^2 = \frac{(8-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(3-5)^2}{5} + \frac{(5-5)^2}{5} = 2,8$$

Dengan α = 0,05, titik 5% atas distribusi χ^2 dengan db = 3 (ingat di sini tidak ada parameter yang ditaksir) adalah 7,815. Oleh karena χ^2 hitungan = 2,8 maka hipotesis tidak ditolak.

Contoh 5.9

Dipunyai data konsumsi listrik dalam kilowatt jam di 92 Dati II sebagai tertuang dalam tabel distribusi frekuensi di bawah. Akan diuji apakah data itu dapat dipandang sebagai sampel random dari suatu populasi normal dengan mean μ dan deviasi standar σ.

Tabel 5.8
Konsumsi Listrik Beberapa Dati II

Konsumsi	Banyak Dati II
500 – 999	2
1000 – 1499	11
1500 – 1999	17
2000 – 2499	25
2500 – 2999	17
3000 – 3499	16
3500 – 3999	3
4000 – 4499	1

Berbeda dengan contoh 5.8, di sini μ dan σ tidak diketahui sehingga harus kita taksir dari data. Nilai μ ditaksir dengan $\overline{x}$ dan σ dengan s, setelah kita hitung diperoleh: $\overline{x} = 2341,89$ dan $s = 740,99$

Selanjutnya kita hitung peluang untuk tiap kelas dan frekuensi harapannya. Kita peroleh sebagai berikut.

Konsumsi	O	Peluang sel	E	$\frac{(O-E)^2}{E}$
500 - 999	2	0,0359	3,30 } 11.69	} 0,15
1000 - 1499	11	0,0912	8,39	
1500 – 1999	17	0,1957	18,00	0,06
2000 – 2499	25	0,2604	23,96	0,05
2500 – 2999	17	0,2301	21,17	0,82
3000 – 3499	16	0,1273	11,71	1,57
3500 – 3999	3 } 4	0,0469	4,31 } 5,46	0,39
4000 – 4499	1	0,0125	1,15	
Jumlah	92	1,00	92	3,04

Cara menghitung:

Pertama-tama kita hitung peluang sel sebagai berikut.

Untuk sel 1 : $Z_1 = \dfrac{1000 - 2341,89}{740,99} = -1,81$

Untuk sel 2 : $Z_2 = \dfrac{1500 - 2341,89}{740,99} = -1,14$

… dan seterusnya.

Dari nilai z_1, z_2, … dan seterusnya kita cari peluang sel melalui tabel kurva normal standar, kita peroleh:

Untuk sel 1 : peluang = 0,0359

Untuk sel 2 : peluang = 0,1271 – 0,0359 = 0,0912

… dan seterusnya.

Maka, E_1 = 0,0359 (0,2) = 3,30

E_2 = 0,0912 (92) = 8,39

…. dan seterusnya.

Sehingga $\chi^2 = \dfrac{(13 - 11,69)^2}{11,69} + ... + \dfrac{(4 - 5,46)^2}{5,46}$

$= 0,15 + \ldots + 0,39 = 3,04.$

Di sini db = 6 – 1 – 2 = 3 (karena ada 2 parameter yang ditaksir).

Dengan tingkat signifikansi α = 0,05, titik 5% atas distribusi χ^2 dengan db = 3 adalah 7,815.

Oleh karena χ^2 hitungan = 3,04 maka hipotesis nol bahwa populasinya berdistribusi normal tidak ditolak.

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

1) Dipunyai frekuensi pengamatan dari 300 pelemparan satu dadu. Apakah data ini membuat kita ragu-ragu bahwa dadu itu seimbang?

Titik dadu	1	2	3	4	5	6	Jumlah
Frekuensi	33	61	49	65	55	37	300

2) Data dalam tabel berikut menunjukkan pengamatan banyak kelahiran pada suatu rumah sakit dalam empat periode kuartal berurutan.

Kuartal	Jan – Mar	Apr – Jun	Jul – Sept	Okt - Des
Banyak kelahiran	110	57	53	80

Diduga telah lahir bayi sebanyak dua kali pada kuartal Januari – Maret dari pada kuartal-kuartal yang lain. Dengan singkat signifikansi $\alpha = 0,10$ ujilah bahwa data ini dengan kuat mendukung pernyataan di atas.

3) *Rumus alternatif untuk* χ^2 *Pearson.* Dengan menjabarkan bentuk kuadrat ruas kanan dari

$$\chi^2 = \sum_{\text{sel}} \frac{(n_i - np_{i0})^2}{np_{i0}}$$

tunjukkan bahwa statistik χ^2 dapat juga ditulis sebagai

$$\chi^2 = \sum_{\text{sel}} \frac{n_i^2}{np_{i0}} - n, \text{ yakni } \sum_{\text{sel}} \frac{O^2}{E} - n$$

4) Data berikut adalah berat (dalam gram) 70 ekor tikus umur 31 hari dalam suatu percobaan.

120	116	94	120	112	112	106	102	118	112
116	98	117	129	130	134	112	122	110	84
106	122	124	112	128	128	106	120	119	106
106	102	140	102	123	124	110	130	116	114
108	110	115	118	117	108	102	125	104	112
132	112	126	122	114	111	98	105	120	106
135	110	139	120	121	126	114	97	116	100

Dengan tingkat signifikansi $\alpha = 0{,}05$ kita ingin menguji hipotesis nol:
H_0: Sampel random itu merupakan observasi pada variabel random berdistribusi normal dengan mean = 110 dan deviasi standar = 10.

5) Dipunyai distribusi frekuensi gaji 171 karyawan di suatu daerah sebagai berikut (dalam ribuan).

Gaji	Frekuensi
100 – 110	9
110 – 120	18
120 – 130	23
130 – 140	23
140 – 150	26
150 – 160	22
160 – 170	18
170 – 180	15
180 – 190	8
190 – 200	4

Apakah data itu dapat dipandang sebagai sampel random dari suatu populasi normal?

RANGKUMAN

Uji kesesuaian χ^2 Pearson:

Data: frekuensi sel pengamatan n_1, n_2, …, n_k dari suatu sampel random berukuran n yang diklasifikasi ke dalam k sel.

Hipotesis nol menyatakan peluang-peluang sel:

$$H_0 : p_1 = p_{10}; ...; p_k = p_{k0}$$

Statistik penguji:

$$\chi^2 = \sum_{sel} \frac{(n_i - np_{i0})^2}{np_{i0}} \quad ; \quad db = k - 1$$

Daerah penolakan: $\chi^2 \geq \chi^2_\alpha$.

Prosedur uji ini berlaku baik untuk uji kesesuaian distribusi multinomial atau distribusi-distribusi yang lain, seperti Poisson, normal.

Hanya saja perlu diperhatikan apakah ada parameter yang tidak diketahui sehingga harus ditaksir dari data. Dalam hal ada parameter yang harus ditaksir dari data maka db-nya berkurang dengan banyak parameter yang ditaksir.

db = (banyak sel) – 1 – (banyak parameter yang ditaksir).

TES FORMATIF 1

Pilihlah satu jawaban yang paling tepat!

1) Untuk menguji kualitas pembentuk angka random, cacah frekuensi untuk tiap angka dicatat dari hasil 500 angka. Konsep kerandoman berarti angka-angka 0; 1; …; 9 berkemungkinan sama. Berdasarkan cacah frekuensi pengamatan di bawah apakah Anda menyangka ada bias dari pembentuk angka random itu?

Angka	0	1	2	3	4	5	6	7	8	9
Frekuensi	43	58	51	59	39	56	45	37	60	52

a. Untuk menjawab pertanyaan di atas kita hitung statistik penguji χ^2, sama dengan ….
   A. 11,00
   B. 13,00
   C 15,00
   D. 18,00

b. Untuk uji kesesuaian kita punyai daerah kritik (α = 5%)= ….
   A. $\chi^2 \geq 14,07$
   B. $\chi^2 \geq 14,68$
   C. $\chi^2 \geq 15,51$
   D $\chi^2 \geq 16,92$

2) Pengamatan 80 keluarga yang masing-masing mempunyai 3 orang anak mengungkapkan distribusi anak laki-laki per keluarga sebagai berikut.

Banyak anak laki-laki	0	1	2	3	Jumlah
Banyak keluarga	19	32	22	7	80

a. Dengan model Bernoulli trial untuk jenis kelamin anak, distribusi peluang banyak anak laki-laki per keluarga haruslah distribusi binomial dengan n = 3 dan p = peluang lahir laki-laki. Dari data di atas, parameter p ditaksir sebagai:

$$\hat{p} = \frac{\text{banyak anak laki-laki dalam 80 keluarga}}{\text{banyak anak seluruhnya dalam 80 keluarga}}$$

yakni, sama dengan ….

   A. 0,2
   B. 0,4
   C. 0,6
   D. 0,8

b. Maka, peluang sel adalah ….

A.

x	p(x)
0	0,216
1	0,432
2	0,288
3	0,064

B.

x	p(x)
0	0,291
1	0,373
2	0,311
3	0,025

C.

x	p(x)
0	0,255
1	0,395
2	0,284
3	0,065

D.

x	p(x)
1	0,222
0	0,444
2	0,333
3	0,001

c. Untuk uji kesesuaian kita hitung statistik penguji χ^2 = ….

A. 4,311
B. 3,564
C. 2,798
D. 1,098

D

. Dengan α = 5% kita punyai daerah kritik ….

A. $\chi^2 \geq 3{,}84$
B. $\chi^2 \geq 4{,}61$
C. $\chi^2 \geq 5{,}99$
D. $\chi^2 \geq 7{,}38$

3) 50 pengukuran keasaman air hujan di daerah KPR adalah sebagai berikut.

3,58	3,80	4,01	4,01	4,05	4,05	4,12	4,18	4,20	4,21
4,27	4,28	4,30	4,32	4,33	4,35	4,35	4,41	4,42	4,45
4,45	4,50	4,50	4,50	4,50	4,51	4,52	4,52	4,52	4,57
4,58	4,60	4,61	4,61	4,62	4,62	4,65	4,70	4,70	4,70
4,70	4,72	4,78	4,78	4,80	5,07	5,20	5,26	5,41	5,48

Kita ingin menguji bahwa data itu merupakan sampel random dari suatu populasi normal dengan mean = 4,55 dan deviasi standar = 0,30

a. Jika K_1 = kuartil pertama distribusi normal itu dan K_3 = kuartil ketiganya maka ….
   A. $K_1 = 4, 349$
      $K_3 = 4{,}751$
   B. $K_1 = 4{,}444$
      $K_2 = 4{,}777$
   C. $K_1 = 4{,}511$
      $K_2 = 4{,}811$
   D. $K_1 = 4{,}521$
      $K_2 = 4{,}871$

b. Kita punyai frekuensi sel pengamatan:

	$(-\infty, K_1]$;	$(K_1, K_2]$;	$(K_2, K_3]$;	(K_3, ∞)
A.	14	13	12	11
B.	15	14	13	8
C.	16	15	10	9
D.	17	12	11	10

c. Untuk uji kesesuaian kita hitung statistik penguji χ^2, sama dengan
   A. 1,12
   B 2, 32
   C 3,76
   D 4,98

d. Dengan $\alpha = 5\%$ daerah kritiknya adalah ….
   A. $\chi^2 \geq 7{,}81$
   B. $\chi^2 \geq 5{,}99$
   C. $\chi^2 \geq 4{,}71$
   D. $\chi^2 \geq 3{,}42$

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 1 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 1.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 2. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 1, terutama bagian yang belum dikuasai.

KEGIATAN BELAJAR 2

Uji Homogenitas dan Independensi

A. UJI HOMOGENITAS

Dari tiap populasi, kita ambil sampel random dengan ukuran yang ditentukan dan mengklasifikasi tiap respons dalam satu dari beberapa kategori. Data ini membentuk tabel kategorik dua arah dengan satu klasifikasi menunjukkan populasi dan yang lain merupakan respons yang kita pelajari. Tujuan kita adalah menguji apakah populasi-populasi itu serupa atau *homogen* dalam peluang sel. Untuk ini, kita akan menentukan apakah proporsi pengamatan dalam tiap kategori respons kira-kira sama untuk semua populasi.

Marilah kita pelajari contoh analisis dengan data, seperti berikut ini.

Contoh 5.10

Untuk membandingkan keefektifan dua jenis makanan (diet) A dan B; 150 bayi diikutkan dalam suatu studi. Diet A diberikan kepada 80 bayi yang dipilih secara random dan diet B diberikan kepada 70 bayi yang lain. Beberapa waktu kemudian, kesehatan setiap bayi diperiksa dan diklasifikasi ke dalam salah satu dari 3 kategori: *istimewa, sedang,* dan *kurang*. Dari frekuensi cacah yang kita catat dalam Tabel 5.8 di bawah kita ingin menguji hipotesis nol bahwa tidak ada perbedaan antara kualitas kedua diet (kedua diet *homogen*).

Tabel 5.9
Kesehatan di Bawah 2 Diet

	Istimewa	Sedang	Kurang	Ukuran Sampel
Diet A	37	24	19	80
Diet B	17	33	20	70
Jumlah	54	57	39	150

Kedua baris Tabel 5.9 adalah hasil dari dua sampel yang independen. Untuk ringkasan deskriptif data ini, sebaiknya kita hitung frekuensi relatif tiap-tiap baris (lihat Tabel 5.9a).

Proporsi atau peluang populasi (yang tidak diketahui) tertuang dalam Tabel 5.9 (b). Ini memungkinkan kita untuk menggambarkan hipotesis nol secara lebih jelas.

Tabel 5.9a
Frekuensi Relatif (dari Tabel 5.9)

	Istimewa	Sedang	Kurang	Ukuran Sampel
Diet A	0,46	0,30	0,24	1
Diet B	0,24	0,47	0,29	1

Tabel 5.9b
Proporsi atau Peluang Populasi

	Istimewa	Sedang	Kurang	Ukuran Sampel
Diet A	p_{A1}	p_{A2}	p_{A3}	1
Diet B	p_{B1}	p_{B2}	p_{B3}	1

Hipotesis nol *tidak ada perbedaan* ekuivalen dengan pernyataan bahwa untuk setiap kategori respons peluangnya sama apakah diet A atau diet B. Dengan demikian, kita rumuskan:

$$H_0 : \ p_{A1} = p_{B1} ; \ p_{A2} = p_{B2} ; \ p_{A3} = p_{B3}.$$

Perhatikan meskipun H_0 menyatakan struktur untuk peluang sel, tetapi tidak memberikan nilai numerik tentang peluang dalam tiap kolom itu.

Jika peluang di bawah H_0 itu kita tulis dengan p_1, p_2, dan p_3 maka frekuensi sel harapan dalam tiap baris akan diperoleh dengan mengalikan peluang ini dengan ukuran sampelnya. Jadi frekuensi harapan dalam baris pertama adalah $80p_1$, $80p_2$, dan $80p_3$, dan dalam baris kedua $70p_1$, $70p_2$, dan $70p_3$. Tetapi ρ_i , i = 1, 2, 3, tidak dinyatakan oleh H_0. Dengan demikian, kita harus menaksir parameter ini supaya dapat memperoleh nilai numerik frekuensi harapan itu.

Jumlah kolom dalam Tabel 5.9, yakni 54, 57, dan 39 adalah cacah frekuensi ketiga kategori respons itu dalam ukuran sampel gabungan 150. Di bawah H_0, peluang taksiran itu adalah

$$\hat{p}_1 = \frac{54}{150} ; \quad \hat{p}_2 = \frac{57}{150} ; \quad \hat{p}_3 = \frac{39}{150} .$$

Menggunakan nilai taksiran ini, frekuensi harapan dalam baris pertama menjadi

$$\frac{(80)(54)}{150}, \frac{(80)(57)}{150}, \text{ dan } \frac{(80)(39)}{150}$$

dan dengan cara serupa untuk baris kedua. Dengan mengingat Tabel 5.9, perhatikan pola yang menarik dalam penghitungan ini:

$$\text{Frekuensi sel harapan} = \frac{\text{(jumlah baris)(jumlah kolom)}}{\text{jumlah keseluruhan}}$$

Tabel 5.10 (a) menyajikan frekuensi pengamatan (O) bersama dengan frekuensi harapan (E) yang ditulis dalam kurung. Tabel 5.9 (b) menghitung ukuran penyimpangan $(O - E)^2/E$ untuk tiap-tiap sel. Menjumlahkan nilai ini untuk seluruh sel kita peroleh nilai statistik χ^2.

Tabel 5.10a
Frekuensi Pengamatan dan Harapan Data dalam Tabel 5.9

	Istimewa	Sedang	Kurang
Diet A	37 (28,8)	24 (30,4)	19 (20,8)
Diet B	17 (25,2)	33 (26,6)	20 (18,2)

Tabel 5.10 (b) Nilai $(O - E)^2/E$

	Istimewa	Sedang	Kurang	Jumlah
Diet A	2.335	1,347	0,156	
Diet B	2,668	1,540	0,178	
				8,224 = χ^2

Untuk menentukan derajat bebasnya, kita gunakan sifat statistik χ^2 yang telah kita pelajari sebelumnya. Nilai χ^2 kita telah dihitung dari dua sampel independen masing-masing dengan derajat bebas $(3 - 1) = 2$ karena di sana ada tiga kategori. Sekarang jumlah db = 2 + 2 = 4 harus dikurangi dengan banyak parameter yang ditaksir. Oleh karena p_1, p_2, dan p_3 harus memenuhi hubungan $p_1 + p_2 + p_3 = 1$ maka sebenarnya hanya ada dua parameter yang tidak diketahui. Dengan demikian, statistik χ^2 kita mempunyai db=4 – 2 = 2.

Dengan db = 2, nilai total titik 5% atas distribusi χ^2 adalah 5,991. Oleh karena χ^2 hitungan = 8,224 lebih besar dari nilai tabel maka hipotesis no. 1 ditolak dengan tingkat signifikansi α=0,05. Jadi, data telah menunjukkan adanya perbedaan yang signifikan antara kualitas kedua diet itu.

Setelah memperoleh χ^2 yang signifikan, sekarang kita harus mempelajari Tabel 5.10 (a) dan 5.10 (b) untuk mencoba menentukan letak sumber signifikansi itu. Kita temukan kontribusi yang besar pada χ^2 datang dari kategori "istimewa" dengan frekuensi relatif 37/80 = 46% bagi diet A dan 17/70 = 24% bagi diet B. Data ini menunjukkan bahwa diet A lebih baik dari diet B.

Termotivasi oleh Contoh 5.10 kita sekarang siap untuk mempelajari prosedur uji χ^2 bagi tabel kategorik r x c yang mempunyai sampel-sampel independen dari r populasi yang diklasifikasikan dalam c kategori respons. Seperti telah kita lihat sebelumnya, frekuensi harapan suatu sel diberikan oleh (jumlah baris x jumlah kolom)/jumlah keseluruhan. Mengenai db x^2 untuk tabel r x c, kita perhatikan bahwa tiap baris menyumbang db = (c – 1) sehingga sumbangan seluruh r baris adalah r (c – 1). Oleh karena sebanyak (c – 1) parameter harus ditaksir maka

$$\begin{aligned} \text{db } \chi^2 &= r(c-1) - (c-1) \\ &= (r-1)(c-1) \\ &= (\text{banyak baris} - 1)(\text{banyak kolom} - 1) \end{aligned}$$

Uji homogenitas χ^2 dalam tabel kategorik:

Hipotesis nol : Dalam tiap kategori respons, peluangnya sama untuk semua populasi

Statistik penguji:

$$\chi^2 = \sum_{\text{sel}} \frac{(O-E)^2}{E}$$

O = frekuensi sel pengamatan

$$E = \frac{(\text{jumlah baris})(\text{jumlah kolom})}{\text{jumlah keseluruhan}}$$

db = (banyak baris –1)(banyak kolom – 1).

Daerah penolakan : $\chi^2 \geq \chi^2_\alpha$

Contoh 5.11

Suatu survei dilakukan untuk menentukan perokok dalam kelompok profesi yang berbeda. Sampel random petani, guru, karyawan (kantor), dan pedagang diinterviu, dan cacah frekuensi pengamatannya dituangkan dalam Tabel 5.11.

Lakukan uji untuk menentukan apakah tingkat kecanduan rokok tampak sama dalam semua keempat kelompok itu.

Tabel 5.11
Tabel Kategorik Perokok versus Profesi

	Perokok	Bukan Perokok	Ukuran Sampel
Petani	32(58,25)	268(241,75)	300
Guru	51(48,54)	199(201,46)	250
Karyawan	67(58,25)	233(241,75)	300
Pedagang	83(67,96)	267(282,04)	350
Jumlah	233	967	1200

Kita tulis proporsi perokok dalam populasi petani, guru, karyawan, dan pedagang masing-masing dengan ρ_1, ρ_2, ρ_3, dan ρ_4. Berdasarkan sampel random dari empat populasi binomial kita ingin menguji hipotesis:

$$H_0 : p_1 = p_2 = p_3 = p_4$$

Frekuensi sel harapan, yang ditunjukkan dengan angka dalam kurung pada Tabel 5.11, dihitung dengan mengalikan jumlah baris dan jumlah kolom dan dibagi dengan 1200. Statistik χ^2 dihitung dalam Tabel 5.12.

Tabel 5.12
Nilai $(O - E)^2$ / E untuk data Tabel 5.10

	Perokok	Bukan Perokok	
Petani	11,83	2,85	
Guru	0,12	0,03	
Karyawan	1,31	0,32	
Pedagang	3,33	0,80	
			20,59 = χ^2
		db = (4 – 1)(2 – 1) = 3	

Dengan derajat bebas = 3, titik 5% atas distribusi χ^2 adalah 7,81 sehingga hipotesis nol ditolak pada $\alpha = 0{,}05$. Akan ditolak juga pada tingkat signifikansi $\alpha = 0{,}01$. Dengan demikian, nilai P-nya kurang dari 0,01.

Dengan memperhatikan Tabel 5.12, kita peroleh bahwa sumbangan besar untuk statistik x^2 datang dari baris pertama. Ini disebabkan karena frekuensi relatif perokok di antara para petani sangat rendah dibandingkan dengan yang lain, sebagaimana dapat kita lihat dari Tabel 5.10.

Contoh 5.12 (*Tabel kategorik 2 × 2*)

Untuk menentukan pengaruh yang mungkin perlakuan kimia pada tingkat persemaian biji, 100 biji yang diberi perlakuan kimia dan 150 biji yang tidak diberi perlakuan kimia ditaburkan. Banyak biji yang disemai dicatat dalam Tabel 5.13. Apakah data memberikan dukungan yang kuat bahwa tingkat persemaian biji yang mendapat perlakuan kimia dan yang tidak berbeda?

Tabel 5.13

	Sesuai	**Tidak sesuai**	**Jumlah**
Mendapat perlakuan	84(86,40)	16(13,60)	100
Tak mendapat perlakuan	132(129,60)	18(20,40)	150
Jumlah	216	34	250

Misalkan, p_1 dan p_2 masing-masing menunjukkan peluang akan semai untuk biji yang mendapatkan perkuan kimia dan yang tidak. Kita ingin menguji hipotesis nol $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$. Untuk uji χ^2 kita hitung frekuensi harapan dengan cara biasa. Ini dituangkan dalam kurung Tabel 5.12. Nilai χ^2 hitungan sama dengan

$$\begin{aligned} x^2 &= 0{,}067 + 0{,}424 + 0{,}044 + 0{,}281 \\ &= 0{,}817 \end{aligned}$$

dengan db = (2 – 1)(2 – 1) = 1.

Nilai 5% atas χ^2 dari tabel distribusi χ^2 dengan db = 1 adalah 3,84. Oleh karena nilai χ^2 hitungan = 0,817 lebih kecil dari 3,84, hipotesis nol tidak ditolak pada tingkat signifikansi α = 0,05. Tingkat persemaian antara biji yang mendapat perlakuan kimia dan yang tidak mendapat perlakuan tidak berbeda secara signifikan.

Cara lain menganalisis tabel kategorik 2 x 2

Mengingat contoh 5.12, kita perhatikan bahwa tabel kategorik 2 x 2, dengan satu tepi tetap, sebenarnya suatu pemaparan sampel-sampel random independen dari dua populasi dikotomi (yakni, 2 kategori). Struktur ini ditunjukkan dalam Tabel 5.14 dengan dua kategori itu kita namakan *sukses* dan *gagal.* Di sini x dan y menunjukkan banyak sukses dalam sampel-sampel yang independen yang masing-masing berukuran n_1 dan n_2 diambil dari populasi 1 dan populasi 2. Misalkan, p_1 dan p_2 menunjukkan peluang akan sukses masing-masing untuk populasi 1 dan populasi 2.

Tabel 5.14
Sampel-sampel independen dari dua populasi dikotomi

	Banyak sukses	**Banyak gagal**	**Ukuran Sampel**
Populasi 1	x	$n_1 - x$	n_1
Populasi 2	y	$n_2 - y$	n_2

Kita akan menguji hipotesis nol $H_0 : p_1 = p_2$. Proporsi sampel

$$\hat{p}_1 = \frac{x}{n_1} \text{ dan } \hat{p}_2 = \frac{y}{n_2}$$

memberikan nilai taksiran untuk p_1 dan p_2. Jika ukuran sampel besar, uji $H_0 : p_1 = p_2$ dapat didasarkan pada statistik penguji.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\text{sesatan standar taksiran}} \approx N(0; 1)$$

Jika peluang sukses bersama di bawah H_0 kita tulis dengan p maka sesatan standar $(\hat{p}_1 - \hat{p}_2)$ dapat kita tulis

$$SS\ (\hat{p}_1 - \hat{p}_2) = \sqrt{p(1-p)}\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Parameter p yang tidak diketahui ditaksir dengan menggabungkan informasi dari dua sampel itu. Proporsi sukses dalam sampel gabungan memberikan

Taksiran gabungan $\hat{p} = \dfrac{x + y}{n_1 + n_2}$

Sesatan standar taksiran $(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Ringkasnya:

Menguji $H_0 : p_1 = p_2$ *dengan sampel besar*

Statistik penguji:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \text{ dengan } \hat{p} = \frac{x + y}{n_1 + n_2}$$

Daerah penolakan tingkat α adalah $|Z| \geq Z_{\alpha/2}$, $Z \leq Z\alpha$ atau $Z \leq -Z\alpha$ bergantung apakah hipotesis alternatifnya $p_1 \neq p_2$, $p_1 > p_2$ atau $p_1 < p_2$. Di sini Z_α menunjukkan titik α atas distribusi N(0 ; 1).

Meskipun statistik penguji Z dan $\chi^2 = \sum_{sel} \frac{(O - E)^2}{E}$ tampak mempunyai bentuk yang sangat berbeda, keduanya ada hubungan yang eksak, yakni

$Z^2 = \chi^2$ (untuk tabel kategorik 2 x 2).

Demikian juga, $Z^2_{\alpha/2}$ sama dengan titik α atas distribusi χ^2 dengan db = 1. Misalnya, dengan α = 0,05 ; $Z^2_{\alpha/2} = (1{,}96)^2 = 3{,}8416$ yang juga merupakan titik 5% atas distribusi χ^2 dengan db = 1 (lihat tabel distribusi normal dan tabel distribusi χ^2). Jadi kedua prosedur uji itu ekuivalen jika hipotesis alternatif dua sisi. Jika hipotesis alternatif itu satu sisi, misalnya H_1 : $p_1 > p_2$ maka hanya uji Z yang sesuai.

Contoh 5.13

Gunakan uji Z untuk data dalam Contoh 5.12. Kita hitung:

$\hat{p}_1 = \dfrac{84}{100} = 0{,}84$; $\hat{p}_1 = \dfrac{132}{150} = 0{,}88$;

taksiran gabungan: $\hat{p}_1 = \dfrac{84 + 132}{100 + 150} = 0,864$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})}\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{-0,04}{\sqrt{(0,864)(0,136)}\ \sqrt{\frac{1}{100} + \frac{1}{150}}} = -0,904$$

Oleh karena $|Z|$ lebih kecil dari $Z_{0,025} = 1,96$, hipotesis nol tidak ditolak pada $\alpha = 0,05$. Perhatikan bahwa $Z^2 = (-0,904)^2 = 0,817$ yang sama dengan hasil $\chi^2 = 0,817$ dalam Contoh 5.12 di atas.

Distribusi normal pendekatan untuk $(\hat{p}_1 - \hat{p}_2)$ memungkinkan kita untuk menghitung interval kepercayaan bagi selisih $(p_1 - p_2)$

> *Interval kepercayaan sampel besar untuk* $(p_1 - p_2)$:
>
> Interval kepercayaan $100(1-\alpha)\%$ pendekatan untuk $(p_1 - p_2)$ adalah
>
> $$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$
>
> Untuk ukuran sampel n_1 dan n_2 besar.

Sebagai contoh, kita sekali lagi kembali ke data Contoh 5.12 dan menghitung:

$\hat{p}_1 = \dfrac{84}{100} = 0,84$ $\qquad\qquad$ $\hat{p}_2 = \dfrac{132}{150} = 0,88$

maka $(\hat{p}_1 - \hat{p}_2) = -0,04$

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_2)}{n_1} + \frac{\hat{\rho}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{(0,84)(0,16)}{100} + \frac{(0,88)(0,12)}{150}} = 0,045$$

Interval kepercayaan 95% pendekatan untuk $(\hat{p}_1 - \hat{p}_2)$ adalah – 0,04 ± (1,96)(0,045) = - 0,04 ± 0,09 atau (~0,13 ; 0,05).

B. UJI INDEPENDENSI

Jika dua sifat diamati untuk setiap elemen suatu sampel random maka data dapat diklasifikasi sekaligus menurut sifat-sifat ini. Dengan demikian, kita peroleh tabel kategorik dua arah yang jumlah kedua tepinya random. Beberapa contoh keadaan ini dapat kita sebutkan: sampel random beberapa orang yang mempunyai pekerjaan tetap dapat diklasifikasi menurut tingkat pendidikan dan jenis kedudukan yang dicapai; mahasiswa diklasifikasi menurut profesi orang tua dan bidang studi yang ditekuni; tanaman berbunga dapat diklasifikasi menurut jenis daunnya dan ukuran bunganya.

Aspek inferensi yang sering dilakukan bagi data yang ditabulasi silang adalah studi tentang apakah kedua sifat itu independen atau apakah satu sifat cenderung berkaitan atau bergantung pada sifat yang lain.

Contoh 5.14

Suatu survei dilakukan untuk mempelajari sikap pemirsa terhadap program TV yang mempertontonkan kekerasan. Sampel random dengan 500 pemirsa dewasa dipilih dan diklasifikasi menurut jenis kelamin dan jawaban atas pertanyaan: Apakah menurut pendapat Anda ada hubungan antara kekerasan di TV dengan tindak kriminal di masyarakat? Hasil klasifikasi ini tertuang dalam Tabel 5.15 berikut ini.

Tabel 5.15
Tabel Kategorik untuk Jenis Kelamin dan Jawaban Pertanyaan Survei

	Ya	Tidak	Tidak tahu	Jumlah
Laki-laki	138	83	64	285
Perempuan	64	67	84	215
Jumlah	202	150	148	500

Membagi frekuensi sel dengan ukuran sampel 500 kita peroleh frekuensi relatif, seperti ditunjukkan dalam Tabel 5.15 (a). Jumlah tepi baris 0,570 dan 0,430 merupakan proporsi sampel laki-laki dan perempuan. Demikian juga, jumlah tepi kolom menunjukkan proporsi sampel ketiga kategori jawaban, yakni masing-masing 0,404; 0,300; dan 0,296.

Tabel 5.15a
Proporsi Pengamatan dalam Tiap Sel

	Ya	Tidak	Tidak tahu	Jumlah
Laki-laki	0,276	0,166	0,128	0,570
Perempuan	0,128	0,134	0,168	0,430
Jumlah	0,404	0,300	0,296	1,00

Bayangkan klasifikasi seluruh populasi. Proporsi populasi yang tidak diketahui (yakni, peluang sel) disajikan dalam Tabel 5.15 (b), dengan indeks L dan P untuk menunjukkan laki-laki dan perempuan, dan 1, 2, dan 3 untuk kategori Ya, Tidak, dan Tidak tahu.

Tabel 5.15b
Peluang Sel

	Ya	Tidak	Tidak Tahu	Peluang Tepi Baris
Laki-laki	p_{L1}	p_{L2}	p_{L3}	p_L
Perempuan	p_{P1}	p_{P2}	p_{P3}	p_P
Peluang tepi kolom	p_1	p_2	$p_{\backslash 3}$	1

Tabel 5.15 (b) adalah analogi populasi Tabel 5.15 (a) yang menunjukkan proporsi sampel. Misalnya,

Peluang sel p_{L1} = P (laki-laki *dan* Ya)
Peluang tepi baris p_L = P_L (laki-laki)
Peluang tepi kolom p_1 = P_1 (Ya)

Kita ingin mempunyai hipotesis nol bahwa kedua klasifikasi independen. Kita tahu bahwa peluang irisan peristiwa yang independen sama dengan hasil kali peluang masing-masing. Jadi independensi dua klasifikasi itu berarti $p_{L1} = p_L . p_1$; $p_{L2} = p_L p_2$ dan seterusnya.

Dengan demikian, hipotesis nol independensi dapat dirumuskan sebagai:

H_0 : Tiap peluang sel sama dengan hasil kali peluang tepi pasangan yang bersangkutan.

Untuk melakukan uji χ^2 kita perlu menentukan frekuensi harapan. Di bawah H_0, frekuensi sel harapan adalah:

$500\,p_L p_1$; $500\,p_L p_2$; $500\,p_L p_3$

$500\,p_P p_1$; $500\,p_P p_2$; $500\,p_P p_3$

Nilai-nilai memuat peluang tepi yang tidak diketahui, yang harus ditaksir dari data. Dari Tabel 5.15, nilai taksiran itu adalah:

$$\hat{p}_L = \frac{285}{500} \quad , \quad \hat{p}_P = \frac{215}{500}$$

$$\hat{p}_1 = \frac{202}{500} \quad , \quad \hat{p}_2 = \frac{150}{500} \quad , \quad \hat{p}_3 = \frac{148}{500}$$

Dengan menggunakan ini, frekuensi harapan untuk tiap sel Tabel 5.15 adalah

$$\frac{\text{(jumlah baris)(jumlah kolom)}}{\text{jumlah seluruhnya}}$$

Misalnya, dalam sel pertama kita punyai

$$500\,\hat{p}_L \hat{p}_1 = 500.\frac{285}{500}.\frac{202}{500} = \frac{285.202}{500} = 115{,}14$$

Tabel 5.15 (c) menyajikan frekuensi sel pengamatan bersama dengan frekuensi harapan yang ditunjukkan dalam kurung. Kuantitas $(O - E)^2/E$ dan statistik χ^2 dihitung dalam Tabel 5.15 (d).

Tabel 5.15c
Frekuensi Sel Pengamatan dan Harapan untuk Data Tabel 5.15

	Ya	Tidak	Tidak tahu
Laki-laki	138 (115,14)	83 (85,50)	64 (84,36)
Perempuan	64 (86,86)	67 (64,50)	84 (63,64)

Tabel 5.15d
Proporsi pengamatan dalam tiap sel

	Ya	Tidak	Tidak tahu	Jumlah
Laki-laki	4,539	0,073	4,914	
Perempuan	6,016	0,097	6,514	
				22,153 = χ^2

Setelah menghitung statistik χ^2, sekarang tinggal menentukan db-nya dengan menggunakan sifat yang kita pelajari di muka. Oleh karena kita hanya mempunyai satu sampel random maka db kita hitung dengan

db = (banyak sel) – 1 – (banyak parameter yang ditaksir)

Oleh karena $p_L + p_p = 1$ dan $p_1 + p_2 + p_3 = 1$, kita sebenarnya hanya menaksir 1 + 2 = 3 parameter. Jadi, db = 6 – 1 – 3 = 2.

Dengan tingkat signifikansi $\alpha =$ 0,05, titik 5% atas distribusi χ^2 adalah 5,991. Oleh karena nilai χ^2 hitungan lebih besar dari tabel ini maka hipotesis nol independen ditolak pada $\alpha = 0{,}05$. Bahkan, hipotesis nol itu akan ditolak meski dengan $\alpha = 0{,}01$.

Dari Tabel 5.15 (d) terlihat bahwa sumbangan yang besar untuk nilai χ^2 datang dari sel-sel di sudut. Lagi pula, membandingkan frekuensi pengamatan dan harapan dalam Tabel 5.15 (c) kita lihat jawaban "Ya" yang banyak dari pemirsa laki-laki.

Dari analisis tabel kategori Contoh 5.14, prosedur untuk uji independensi dalam kategori r x c segera tampak. Sebenarnya, ini sama saja dengan uji homogenitas yang telah kita pelajari. Frekuensi sel harapan ditentukan dengan cara yang sama, yakni

$$\text{Frekuensi sel harapan} = \frac{\text{(jumlah baris)(jumlah kolom)}}{\text{jumlah keseluruhan}}$$

dan statistik pengujinya sama dengan

$$\chi^2 = \sum_{sel} \frac{(O-E)^2}{E}$$

Mengenai derajat bebas (db) χ^2 dalam kasus ini, mula-mula kita punyai db = (rc – 1) karena ada rc sel tempat kita mengklasifikasi satu variabel random. Dari sini kita harus mengurangi dengan banyak parameter yang ditaksir. Ini adalah (r – 1) + (c – 1) karena ada (r – 1) parameter di antara peluang tepi baris dan (c –1) parameter di antara peluang tepi kolom.

Maka db χ^2 sama dengan:

$$\begin{aligned} Db &= rc - 1 - (r - 1) - (c - 1) \\ &= rc - r - c + 1 \\ &= (r - 1)(c - 1) \\ &= (\text{banyak baris} - 1)(\text{banyak kolom} - 1) \end{aligned}$$

yang identik dengan (db) x^2 untuk menguji homogenitas. Ringkasnya, statistik penguji χ^2, db-nya, dan daerah penolakan untuk uji independensi sama seperti untuk uji homogenitas. Hanya pernyataan tentang hipotesis nol-nya yang berbeda antara dua keadaan itu.

Hipotesis nol independensi H0 : Tiap peluang sel sama dengan hasil kali peluang tepi baris dan peluang tepi kolom yang bersangkutan.

Catatan:

Jika uji χ^2 menghasilkan kesimpulan menolak hipotesis nol independensi, kita simpulkan bahwa data memberikan kenyataan adanya *keterkaitan statistik* antara 2 sifat yang kita pelajari. Tetapi, kita harus tidak membuat interpretasi yang gegabah bahwa sifat-sifat itu berhubungan langsung. Penegasan hubungan sebab-akibat harus diambil dari pikiran sehat yang fakta statistik tidak boleh menggantikannya.

Dua sifat mungkin kelihatan sangat berhubungan karena sama-sama dipengaruhi faktor ketiga yang tidak termasuk dalam studi. Dalam hal seperti ini, dependensi (ketergantungan) itu dinamakan *dependensi palsu.*

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

1) Banyak pencemar udara dari industri mempunyai pengaruh yang merugikan pada tanaman. Sulfur dioksida menyebabkan kerusakan daun dalam banyak pohon-pohon yang sensitif. Dalam suatu studi tentang pengaruh konsentrasi sulfur dioksida di udara pada tiga jenis sayuran, 40 tanaman dari tiap jenis dikenai pencemaran udara dalam pengendalian kondisi rumah hijau. Diperoleh faktor sebagai berikut:

	Kerusakan daun		Jumlah
	Parah	Tidak Parah	
Sawi	32	8	40
Bayam	28	12	40
Tomat	19	21	40
Jumlah	79	41	120

Analisislah data tersebut guna menentukan apakah kerusakan daun yang parah sepadan untuk tiga jenis sayuran itu. Khususnya, lakukan:

a. perumusan hipotesis nol;
b. uji hipotesis nol dengan $\alpha = 0{,}10$;
c. hitung interval kepercayaan 95% individual ketiga jenis itu dan gambarkan pada kertas grafik.

2) Suatu studi dilakukan untuk membandingkan tingkat ketahanan tubuh terhadap suatu penyakit tertentu di antara anak laki-laki dan perempuan dalam kelompok umur 5 – 9 tahun. Di antara 113 anak laki-laki yang diuji, terdapat 34 anak yang mempunyai ketahanan tubuh; dan di antara 139 anak perempuan yang diuji, terdapat 54 anak yang mempunyai ketahanan tubuh. Apakah data menunjukkan kenyataan mendukung tingkat ketahanan tubuh anak perempuan lebih tinggi dari pada anak laki-laki? Gunakan uji Z dengan $\alpha = 5\%$.
Selanjutnya lakukan hal-hal berikut.

a. Tuliskan data tersebut dalam bentuk tabel kategorik 2×2.
b. Lakukan uji χ^2 dengan $\alpha = 5\%$

c. Bagaimana hubungan antara uji χ^2 dan uji Z di atas?

3) Sampel dengan 408 anak yang diadopsi pada waktu masih bayi diklasifikasi menurut kebiasaan menggunakan tangan mereka dan orang tua angkat mereka.

	Kebiasaan anak		Jumlah
	Tidak kidal	Kidal	
Keduanya tidak kidal	307	48	355
Paling sedikit satu tangan yang kidal	47	6	53
Jumlah	354	54	408

Ujilah hipotesis nol bahwa kedua faktor itu independen. Gunakan α=5%.

4) Suatu survei dilakukan untuk mempelajari sikap mahasiswa terhadap peraturan Rektor yang baru. Suatu sampel random dengan 1250 orang mahasiswa dipilih dan diklasifikasi menurut jenis kelamin dan sikapnya. Diperoleh fakta sebagai berikut:

	Sikap		
	Setuju	Tidak Setuju	Blangko
Laki-laki	378	237	26
Perempuan	438	146	25

Apakah data survei menunjukkan perbedaan yang signifikan antara sikap mahasiswa laki-laki dan mahasiswa perempuan?

RANGKUMAN

Uji homogenitas tabel kategorik r x c

Data : Sampel-sampel random independen dari r populasi, masing-masing sampel diklasifikasi dalam c kategori respons.

Hipotesis nol : Dalam tiap kategori respons, peluangnya sama untuk semua populasi

Statistik penguji : $\chi^2 = \sum_{\text{sel}} \frac{(O-E)^2}{E}$; db = (r − 1)(c − 1)

di mana untuk setiap sel:
O = frekuensi sel pengamatan

$$E = \frac{\text{jumlah baris x jumlah kolom}}{\text{jumlah keseluruhan}}$$

Daerah penolakan : $\chi^2 \geq \chi^2_\alpha$

Membandingkan dua proporsi binomial – tabel kategorik 2 x 2

Data : X = banyak sukses dalam n_1 trial dengan $P(S) = p_1$
Y = banyak sukses dalam n_2 trial dengan $P(S) = p_2$

a. Untuk menguji $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$ dapat kita gunakan uji χ^2 atau uji Z dengan

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1\text{-}\hat{p})}\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

dengan daerah kritik:

$$|Z| \geq Z_{\alpha/2} \text{ dan } \hat{p}_1 = \frac{x}{n_1},\ \hat{p}_2 = \frac{y}{n_2},\ \hat{p} = \frac{x + y}{n_1 + n_2}$$

Kedua uji ekuivalen dengan $Z^2 = \chi^2$

b. Untuk menguji $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$ gunakan uji Z dengan daerah
kritik $Z \geq Z\alpha$. Uji χ^2 tidak dapat digunakan di sini.

c. Interval kepercayaan 100 (1 - α)% untuk $(p_1 - p_2)$ adalah

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\partial/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Uji independensi dalam tabel kategorik r x c

Data : Suatu sampel random berukuran n diklasifikasi menurut dua sifat, satu dengan r kategori dan yang lain dengan c kategori.

Hipotesis nol : Kedua sifat itu independen, yakni tiap peluang sel adalah hasil kali peluang tepi baris dan kolom.

Statistik pengujian daerah penolakannya sama seperti dalam uji homogenitas.

Batasan

Semua prosedur inferensi di sini memerlukan sampel-sampel besar. Uji χ^2 sesuai untuk dipakai jika tidak ada frekuensi sel harapan yang terlalu kecil (lebih besar atau sama dengan 5 adalah yang biasanya disyaratkan).

TES FORMATIF 2

Pilihlah satu jawaban yang paling tepat!

1) Suatu sampel dengan 100 orang perempuan dari kelompok etnik A dan sampel lain dengan 100 orang perempuan dari etnik B masing-masing ditanya: "Apakah Anda menikah sebelum umur 19 tahun?" diperoleh data sebagai berikut:

	A	B
Ya	62	29
Tidak	38	71

a. Untuk uji homogenitas kita hitung frekuensi sel harapan, kita peroleh ….

A.

	A	B
Ya	45,5	45,5
Tidak	54,5	54,5

B.

	A	B
Ya	47,5	47,5
Tidak	42,5	42,5

C.

	A	B
Ya	46,5	43,5
Tidak	46,5	43,5

D.

	A	B
Ya	48,5	41,5
Tidak	41,5	48,5

b. Untuk uji homogenitas kita hitung statistik penguji χ^2 maka diperoleh ….

A. 15,12

B. 17,64
C. 19,31
D. 21,96

c. Dengan $\alpha = 5\%$ daerah kritik uji itu adalah ….
A. $\chi^2 \geq 7{,}38$
B. $\chi^2 \geq 5{,}99$
C. $\chi^2 \geq 3{,}84$
D. $\chi^2 \geq 2{,}61$

d. Untuk uji homogenitas dapat juga kita hitung statistik penguji Z, kita peroleh ….
A. 1,96
B. 2,32
C. 3,67
D. 4,92

e. Kita hitung interval kepercayaan 95% untuk (p_1 - p_2), dengan p_1 = peluang yang ada pada etnik A dan p_2 = peluang ya pada etnik B= ….
A. (0,20 ; 0,46)
B. (0,11 ; 0,55)
C. (0,15 ; 0,39)
D. (0,17 ; 0,40) .

2) Sampel random dengan 350 orang mahasiswa suatu universitas memberikan fakta tentang hidup mereka terhadap peraturan baru di universitas itu.

	Setuju	**Menolak**	**Tak Berpendapat**	**Jumlah**
Laki-laki	93	21	72	186
Perempuan	55	30	79	164
Jumlah	148	51	151	350

Untuk menguji bahwa kedua sifat (sikap dan jenis kelamin) independen kita hitung frekuensi sel harapan sebagai berikut.

	Setuju	Menolak	Tak berpendapat
Laki-laki	E_{11}	E_{12}	E_{13}
Perempuan	E_{21}	E_{22}	E_{23}

a. E_{11} sama dengan ….
   A. 76,86
   B. 77,77
   C. 78,65
   D. 79,11

b. E_{23} sama dengan ….
   A. 70,75
   B. 71,17
   C. 72,25
   D. 72,95

c. E_{22} sama dengan ….
   A. 21,12
   B. 23,90
   C. 24,14
   D. 24,99

d. Kita hitung statistik penguji x^2, sama dengan ….
   A. 8,33
   B. 10,33
   C. 12,33
   D. 14,33

e. Dengan $\alpha = 5\%$ daerah kritiknya adalah ….
   A. $\chi^2\overline{Y} \geq 1,89$
   B. $\chi^2 \geq 2,71$
   C. $\chi^2 \geq 3,84$
   D. $\chi^2 \geq 5,99$

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 2 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 2.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan modul berikutnya. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 2, terutama bagian yang belum dikuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

1) a. B
   b. D
2) a. B
   b. A
   c. D
   d. C
3) a. A
   b. B
   c. B
   d. A

Tes Formatif 2

1) a. A
   b. D
   c. C
   d. B
   e. A
2) a. C
   b. B
   c. B
   d. B
   e. D

Daftar Pustaka

Battachar, G.K. and R.A. Johnson. (1977). *Statistics Concepts and Methods.* New York: John Willey.

Freud. J. (1979). *Modern Elementary Statistics.* Prentice Hall.

Kooros, A. (1965). *Elements of Mathematical Economics.* Houghton Miffing Company, Boston.

Pfeffenberber, R.C. and J.H. Peterson. (1977). *Statistical Methods for Business and Economics.* Illinois: Richard D. Irwin.

Robbins, H. and J.V. Ryzin. (1975). *Introduction to Statistics.* Science Research Associates. Inc.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.