

Wykład 12: Tablice wielodzielcze

<i>Drosophila melanogaster</i> Krzyżówka wsteczna (CcNn i ccnn)		Kolor oczu	
		czerwone	fioletowe
Rozmiar skrzydła	normalne	39	11
	mniejsze	18	32

Zródło: http://pl.wikipedia.org/wiki/Plik:Drosophila_melanogaster1.jpg



Badanie zależności zmiennych kolumnowej i wierszowej:

- Tutaj zmiennymi badanymi są:
- H_0 : zmienne są niezależne
- Obok obserwowanych licznosci, ozn. O,
- Rozważamy licznosci oczekiwane przy H_0 :

$$E := \frac{\text{suma wiersza} \times \text{suma kolumny}}{\text{liczba wszystkich obserwacji}}$$

Uzupełniamy tabelę wartościami oczekiwanymi przy H_0 :

		Kolor oczu		Suma
		czerwone	fioletowe	
Kształt skrzydła	normalne	39 ()	11 ()	50
	mniejsze	18 ()	32 ()	50
Suma		57	43	100

Czy w badanej populacji kolor oczu i kształt skrzydła są zmiennymi niezależnymi?

- H_0 : kolor oczu i rozmiar skrzydła są niezależne
 - H_A : kolor oczu i rozmiar skrzydła są zależne
- Zastosujemy test chi-kwadrat dla niezależności:
 $X^2_s = \sum (O-E)^2/E$ ma przy H_0 tzw. rozkład χ^2_1 .
- Testujemy na poziomie $\alpha = 0.05$; odrzucamy gdy $X^2_s > 3.84 = X^2_{\text{krytyczne}}$
 - Tutaj $X^2_s = \dots$
 - Wniosek:...

Rozkłady brzegowe i warunkowe:

- Opisz rozkład koloru oczu w danych:
- Opisz rozkład kształtu skrzydeł w danych:
- Opisz rozkład koloru oczu w danych dla normalnego kształtu skrzydeł:

Dyskusja rozkładów brzegowych i warunkowych w populacji:

- $p_1 := \Pr(\text{czerwone oczy} \mid \text{normalne skrzydła})$,
- $p_2 := \Pr(\text{czerwone oczy} \mid \text{mniejsze skrzydła})$
- H_0 : (kolor oczu i rozmiar skrzydła są niezależne) jest równoważna równości $p_1 = p_2$
- H_A : (kolor oczu i rozmiar skrzydła są zmiennymi zależnymi) oznacza, że $p_1 \neq p_2$
- Estymatorami p_1 i p_2 są:
 $\hat{p}_1$ $\hat{p}_2$
- Omówimy później przedziały ufności dla p_1, p_2 .

Uwagi:

- Nie możemy powiedzieć, że krótsze skrzydła powodują, że muszka ma fioletowe oczy. Prawidłowy wniosek to obserwacja, że kolor oczu i kształt skrzydła są zmiennymi zależnymi, albo że u muszek z normalnymi skrzydłami częściej występują czerwone oczy niż u muszek z mniejszymi skrzydłami.
- Nie możemy formułować wniosku przyczynowego, także dlatego, że nie kontrolujemy analizowanych zmiennych, a jedynie je obserwujemy.
- W tym wypadku zależność wynika z faktu, że geny determinujące kształt oczu i rozmiar skrzydła leżą na jednym chromosomie.

Rozkład chi-kwadrat

Definicja:

Jeżeli $Z_1, \dots, Z_d$ są niezależne

i mają rozkład $N(0,1)$,

to $X = Z_1^2 + \dots + Z_d^2$ ma rozkład χ^2_d .

Funkcje gęstości rozkładów chi-kwadrat z
df=2 and 4 stopniami swobody:

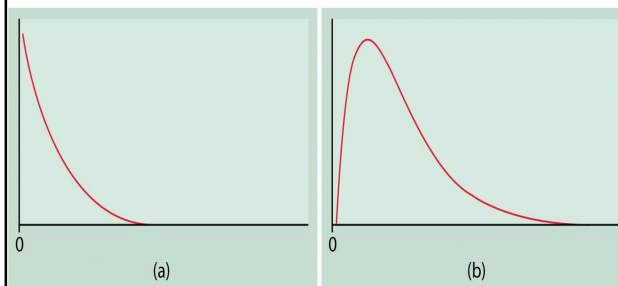


TABLE F χ^2 distribution critical values

df	Tail probability p									
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	22.91	25.19	27.11
11	13.70	14.63	15.72	17.28	19.68	21.92	22.62	24.72	26.76	28.73
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34
80	88.15	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3

Tablice wieloznaczne: $r \times k$

- r rzędów, k kolumn: $r \times k$
- Analiza analogiczna do tablic 2×2 .
- Przykład: 3×4 ($r = 3$; $k = 4$)

		Kolor włosów				Suma
		Brązowe	Czarne	Jasne	Rude	
Kolor oczu	Brązo- we	438 (331.7)	288 (154.1)	115 (356.5)	16 (14.6)	857
	Szare/ Zielone	1387 (1212.3)	746 (563.3)	946 (1303.0)	53 (53.4)	3132
	Niebies- kie	807 (1088.0)	189 (505.6)	1768 (1169.5)	47 (48.0)	2811
Suma		2632	1223	2829	116	6800

Pytanie naukowe: Czy kolor oczu i włosów są zmiennymi zależnymi?

- H_0 : Kolor włosów i kolor oczu to zmienne niezależne
- H_A : Kolor oczu i kolor włosów to zmienne zależne
- Wykonujemy test niezależności chi-kwadrat
- $df = (r-1)(k-1) = \dots$
- $X^2 = \sum (O-E)^2/E$ ma przy H_0 rozkład χ^2_6 .

- Testujemy na poziomie $\alpha = .0005$.
- Wartość krytyczna $\chi^2_6 = \dots$
- $X^2_s = \dots$
- Wniosek...
- Tablica wartości krytycznych z książki "Introduction to the Practice of Statistics", D.S. Moore, G. P. McCabe

- Estymator dla $Pr(\text{Oczy niebieskie}) = \dots$
- Estymator dla $Pr(\text{Oczy niebieskie} | \text{włosy brązowe}) = \dots$
- Estymator dla $Pr(\text{Oczy niebieskie} | \text{czarne włosy}) = \dots$
- Estymator dla $Pr(\text{Oczy niebieskie} | \text{jasne włosy}) = \dots$
- Estymator dla $Pr(\text{Oczy niebieskie} | \text{rude włosy}) = \dots$

Uwagi:

- Testowanie niezależności odpowiada testowaniu, że odpowiednie p-stwa warunkowe są te same w każdej klasie.
- Gdy testujemy niezależność w dużych tabelach, to na ogół nie zapisujemy H_0 za pomocą prawdopodobieństw warunkowych.
- Założenia:
 - ✓ Próby proste z odpowiednich populacji (reprezentatywne, niezależne, losowe)
 - ✓ "E" w każdej komórce musi być ≥ 5

Przykład (Aspiryna):

- 21,996 amerykańskich mężczyzn lekarzy w średnim wieku
- Połowa brała aspirynę, połowa-placebo.
- Po trzech latach, 139 spośród biorących aspirynę i 239 spośród biorących placebo miało atak serca.
- Zbadaj zależność pomiędzy braniem aspiryny i występowaniem ataku serca.

Dyskusja

- W artykule prasowym czytamy, że 80% pieszych będących ofiarami nocnych wypadków samochodowych nosiło ciemne ubrania, a 20% jasne ubrania.
- Wyciągnięto wniosek, że w nocy bezpieczniej jest nosić jasne ubrania.
- Czy przeprowadzone badania upoważniają do takiej konkluzji?

Przedziały ufności dla różnicy między p-stwami warunkowymi

- W tabelach 2x2, wyrażamy H_0 jako $p_1 = p_2$
- Przykład z aspiryną:
- $p_1 = \text{Pr}(\text{atak serca} \mid \text{aspiryna})$,
 $p_2 = \text{Pr}(\text{atak serca} \mid \text{placebo})$.

$$SE_1 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}, SE_2 = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, SE = \sqrt{SE_1^2 + SE_2^2}$$

- Przybliżony 95% PU dla $p_1 - p_2$ wynosi

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96 SE$$

- W przykładzie z aspiryną:

$$\hat{p}_1 = \quad \hat{p}_2 =$$

$$SE =$$

- PU dla $p_1 - p_2$ wynosi:...

- Mamy 95% pewności, że częstość ataku serca przy regularnym pobieraniu (odpowiedniej dozy aspiryny) zmniejsza się od ... do ... punktów procentowych (na przestrzeni 3 lat).
- W ogólności do konstrukcji przedziałów ufności na poziomie $(1-\alpha)$ stosujemy $Z_{\alpha/2}$ (zamiast 1.96).

Uwagi:

- W teście chi-kwadrat w tablicach wielodzielczych i dla przedziałów ufności dla różnicy proporcji stosujemy przybliżenie rozkładem normalnym. Obowiązują zwykłe warunki sformułowane wcześniej.
- Ważne jest zwłaszcza, aby próba była reprezentacyjna i losowa.

Dokładny test Fishera

- Stosujemy dla małych rozmiarów prób
- Przykład: ECMO
- ECMO to ``nowa'' procedura służąca ratowaniu noworodków cierpiących na poważne zaburzenia pracy układu oddechowego.
- CMT – konwencjonalna terapia

		Zabieg		
Wynik		CMT	ECMO	Suma
	Zgon	4	1	5
	Przeżycie	6	28	34
	Suma	10	29	39

- H_0 : wynik nie zależy od zabiegu
- Znajdziemy warunkowe prawdopodobieństwo zaobserwowanych wyników przy ustalonych ``sumach'' w rzędach i kolumnach (przy H_0).
- Przypomnijmy symbol Newtona - $\binom{n}{k}$
- Na tyle sposobów można wybrać zbiór k elementowy ze zbioru n elementowego

- Na ile sposobów dokładnie 4 dzieci spośród 5 z tych które „miały” umrzeć mogło przypadkowo zostać przyporządkowanych do grupy CMT:...
- Na ile sposobów dokładnie 6 dzieci spośród 34 z tych które „miały” przeżyć mogło przypadkowo zostać przyporządkowanych do grupy CMT:...
- Na ile sposobów 10 dzieci spośród 39 mogło przypadkowo zostać przyporządkowanych do grupy CMT:...

- H_A : ECMO jest lepsza niż CMT
- Przypadki bardziej ekstremalne w kierunku alternatywy
liczba śmierci = CMT:4, ECMO:1 $\rightarrow$ CMT:5,
ECMO:0
- P-wartość =...
- Wniosek:...