

Testowanie hipotez statystycznych.

Krzysztof Topolski

Bioinformatyka

Wykład 6

Wrocław, 5 listopada 2012

Temat.

Weryfikacja hipotez statystycznych dotyczących wariacji populacji o rozkładzie normalnym.

Test χ^2 (chi-kwadrat).

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, Testujemy hipotezę

$$H_0 : \sigma^2 = \sigma_0^2$$

przeciwko jednej z następujących hipotez alternatywnych

$$H_1 : \sigma^2 < \sigma_0^2,$$

$$H_2 : \sigma^2 > \sigma_0^2,$$

$$H_3 : \sigma^2 \neq \sigma_0^2.$$

Test χ^2 (chi-kwadrat).

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, Testujemy hipotezę

$$H_0 : \sigma^2 = \sigma_0^2$$

przeciwko jednej z następujących hipotez alternatywnych

$$H_1 : \sigma^2 < \sigma_0^2,$$

$$H_2 : \sigma^2 > \sigma_0^2,$$

$$H_3 : \sigma^2 \neq \sigma_0^2.$$

Test χ^2 (chi-kwadrat).

Ponieważ wariancja jest nieznana więc przy konstrukcji statystyki testowej użyjemy nieobciążonego estymatora wariancji S^2

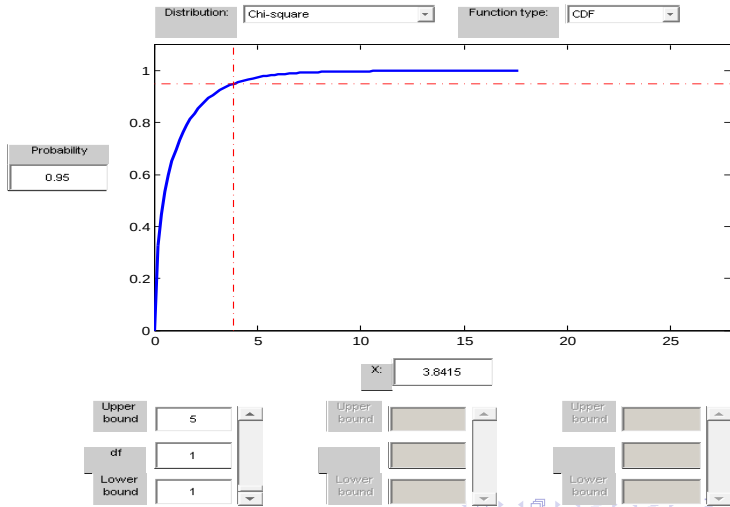
$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Jako statystyki testowej użyjemy następującej statystyki

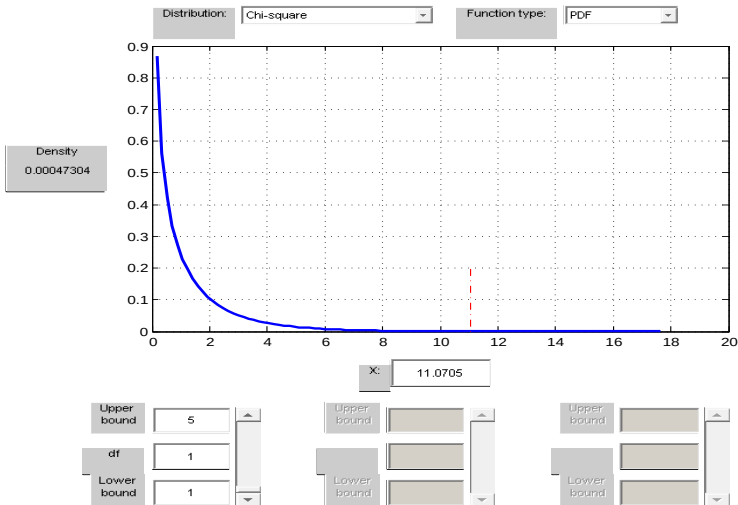
$$T = \frac{(n-1)S_n^2}{\sigma_0^2}.$$

Przy założeniu hipotezy zerowej H_0 statystyka T ma rozkład χ^2 z $(n-1)$ stopniami swobody.

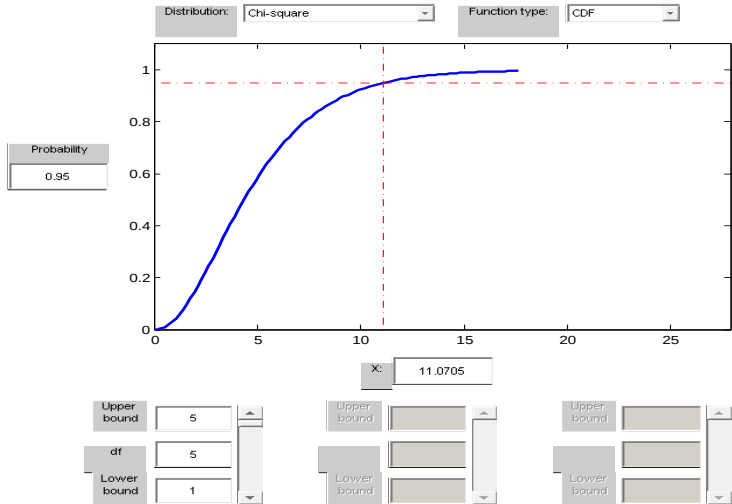
Dystrybuanta rozkładu χ^2 z jednym stopniem swobody.



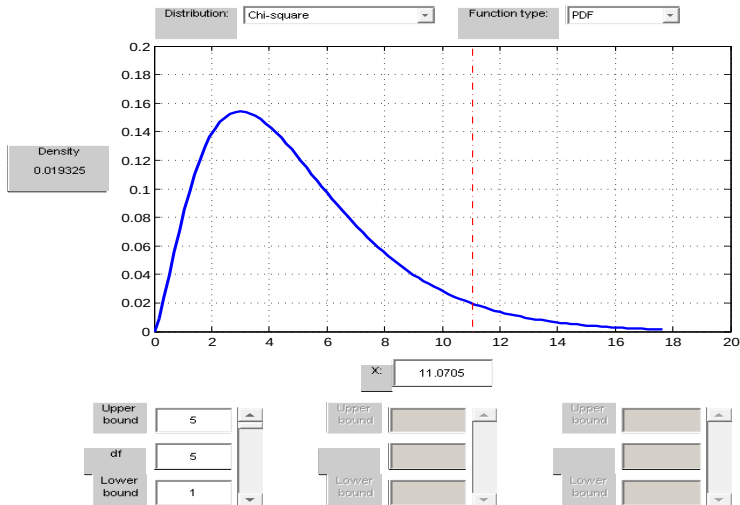
Gęstość rozkładu χ^2 z jednym stopniem swobody.



Dystrybuanta rozkładu χ^2 z 5 stopniami swobody.



Gęstość rozkładu χ^2 z 5 stopniami swobody.



Test χ^2 (chi-kwadrat).

Niech

$\chi^2_{(n-1)}(\alpha/2)$ oznacza kwantyl rzędu $\alpha/2$

natomiast

$\chi^2_{(n-1)}(1 - \alpha/2)$ oznacza kwantyly rzędu $(1 - \alpha/2)$,

rozkładu χ^2 z $(n - 1)$ stopniami swobody.

Na poziomie istotności α zbiory krytyczne testów dla poszczególnych alternatyw mają odpowiednio postać:

- 1) $C = [0, \chi^2_{(n-1)}(\alpha)]$ – dla alternatywy H_1
- 2) $C = [\chi^2_{(n-1)}(1 - \alpha), \infty)$ – dla alternatywy H_2
- 3) $C = [0, \chi^2_{(n-1)}(\alpha/2)] \cup [\chi^2_{(n-1)}(1 - \alpha/2), \infty)$
– dla alternatywy H_3

Test χ^2 (chi-kwadrat).

Postać obszaru krytycznego w przypadku hipotez jednostronnych wynika z następującego rozumowania. Jeżeli hipotezę dwustronną zastąpimy hipotezą jednostronną,

$$H_1 : \sigma^2 < \sigma_0^2,$$

to zbiór krytyczny ma postać

$$C = [0, \chi_{(n-1)}^2(\alpha)].$$

Możemy napisać

$$T = \frac{(n-1)S_n^2}{\sigma_0^2} = \left(\frac{(n-1)S_n^2}{\sigma^2} \right) \left(\frac{\sigma^2}{\sigma_0^2} \right).$$

Porównanie wariancji w dwóch niezależnych populacjach o rozkładach normalnych.

Test F (test Fishera)

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, a $Y_1, \dots, Y_m$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_2, \sigma_2^2)$. Zakładamy, że σ_1^2 oraz σ_2^2 są **nieznane**. Testujemy hipotezę

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

przeciwko jednej z następujących hipotez alternatywnych

$$H_1 : \sigma_X^2 < \sigma_Y^2,$$

$$H_2 : \sigma_X^2 > \sigma_Y^2,$$

$$H_3 : \sigma_X^2 \neq \sigma_Y^2.$$

Test F (test Fishera)

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, a $Y_1, \dots, Y_m$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_2, \sigma_2^2)$. Zakładamy, że σ_1^2 oraz σ_2^2 są **nieznane**. Testujemy hipotezę

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

przeciwko jednej z następujących hipotez alternatywnych

$$H_1 : \sigma_X^2 < \sigma_Y^2,$$

$$H_2 : \sigma_X^2 > \sigma_Y^2,$$

$$H_3 : \sigma_X^2 \neq \sigma_Y^2.$$

Test F (test Fishera)

Przy konstrukcji statystyki testowej użyjemy nieobciążonych estymatorów wariancji S_X^2 i S_Y^2

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

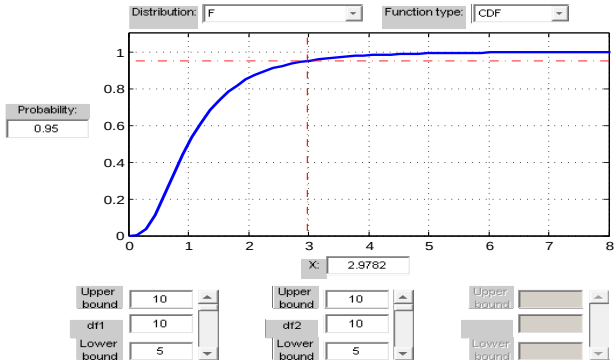
$$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2.$$

Jako statystyki testowej użyjemy

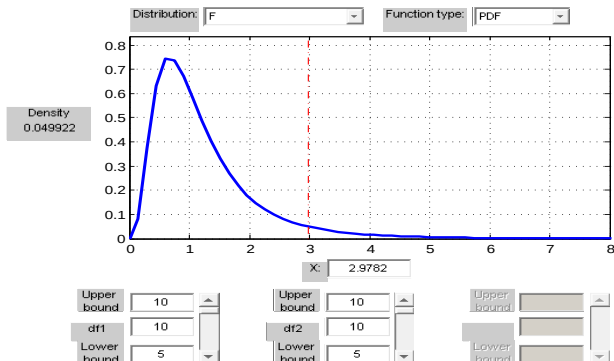
$$F = \frac{S_X^2}{S_Y^2}.$$

Gdy hipoteza H_0 jest prawdziwa to statystyka F ma rozkład F –Snedecora z $(n - 1)$ i $(m - 1)$ stopniami swobody.

Dystrybuanta rozkładu F z 10 stopniami swobody.



Gęstość rozkładu F z 10 stopniami swobody.



Test F (test Fishera)

Niech

$f_{n-1,m-1}(\alpha/2)$ oznacza kwantyl rzędu $\alpha/2$

natomiast

$f_{n-1,m-1}(1 - \alpha/2)$ oznacza kwantyl rzędu $(1 - \alpha/2)$,

rozkładu F –Snedecora z $(n - 1)$ i $(m - 1)$ stopniami swobody.

Na poziomie istotności α zbiory krytyczne testów dla poszczególnych alternatyw mają odpowiednio postać:

- 1) $C = [0, f_{n-1,m-1}(\alpha)]$ – dla alternatywy H_1
- 2) $C = [f_{n-1,m-1}(1 - \alpha), \infty)$ – dla alternatywy H_2
- 3) $C = [0, f_{n-1,m-1}(\alpha/2)] \cup [f_{n-1,m-1}(1 - \alpha/2), \infty)$
– dla alternatywy H_3

Porównanie średnich w dwóch rozkładach normalnych (*brak założenia o równości wariancji*).

Test Cochran-Coxa.

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, a $Y_1, \dots, Y_m$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_2, \sigma_2^2)$. Zakładamy, że σ_1^2 oraz σ_2^2 są **nieznane**. Testujemy hipotezę

$$H_0 : \mu_1 = \mu_2$$

przeciwko hipotezie alternatywnej

$$H_1 : \mu_1 \neq \mu_2.$$

Test Cochran-Coxa.

Niech $X_1, \dots, X_n$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_1, \sigma_1^2)$, a $Y_1, \dots, Y_m$ będzie próbą statystyczną z rozkładu normalnego $N(\mu_2, \sigma_2^2)$. Zakładamy, że σ_1^2 oraz σ_2^2 są **nieznane**.
Testujemy hipotezę

$$H_0 : \mu_1 = \mu_2$$

przeciwko hipotezie alternatywnej

$$H_1 : \mu_1 \neq \mu_2.$$

Test Cochran-Coxa.

Ponieważ wariancje są nieznane więc przy konstrukcji statystyki testowej użyjemy nieobciążonych estymatorów wariancji S_X^2 i S_Y^2

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

$$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2.$$

Jako statystyki testowej użyjemy statystykę Cochran-Coxa

$$C = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

gdzie $\overline{X}_n$ i $\overline{Y}_m$ oznaczają odpowiednie średnie próbkowe, a S_X^2 i S_Y^2 oznaczają wariancje próbkowe.

Gdy hipoteza H_0 jest prawdziwa to statystyka C ma rozkład zależny od nieznanego stosunku $\frac{\sigma_1^2}{\sigma_2^2}$.

Test Cochran-Coxa.

Dla zadanych liczebnościach prób n i m można znaleźć przybliżoną wartość $c_{n,m}(\alpha)$ kwantyla rzędu α rozkładu statystyki C , a mianowicie

$$c_{n,m}(\alpha) \approx \frac{\frac{s_1^2}{n} t_{n-1}(\alpha) + \frac{s_2^2}{m} t_{m-1}(\alpha)}{\frac{s_1^2}{n} + \frac{s_2^2}{m}},$$

gdzie $t_n(\alpha)$ oznacza kwantyl rzędu α rozkładu t -Studenta o n stopniach swobody. Na poziomie istotności α zbiór krytyczny ma postać:

$$K = \{c : |c| > c_{n,m}(1 - \alpha/2)\},$$

gdzie c jest wyliczoną na podstawie danych wartością statystyki C .