#### Prawdopodobieństwo całkowite i warunkowe

Prawdopodobieństwo całkowite. Niech będzie dana przestrzeń probabilistyczna  $(\Omega, \Sigma, P)$  oraz zdarzenia  $A_1, A_2, A_n \in \Sigma$  spełniająca warunki:  $P(A_i) > 0$  dla każdego  $i = 1, ..., n; A_i \cap A_j = \emptyset$  dla wszystkich  $i \neq j; A_1 \cup ... A_n = \Omega$ 

Prawdopodobieństwo warunkowe:  $P(B|A) = \frac{P(B \cap A)}{P(A)}$ . Wtedy dla każdego zdarzenia  $B \in \Sigma$  zachodzi następująca równość:  $P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)$ 

Wzór Bayesa:  $P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}.$  Niezależność zdarzeń:  $P(A\cap B) = P(A)\cdot P(B), \ P(A_{k_1}\cap\ldots\cap A_{k_r}) = P(A_{k_1})\cdot\ldots\cdot P(A_{k_r})$ 

# Wartość oczekiwana i wariancja

Wartość oczekiwana dla rozkładu dyskretnego:  $m = E(X) = \sum_{i=1}^{n} x_i p_i$ , ciągłego:  $m = E(X) = \int_{-\infty}^{\infty} x f(x) dx$ 

Wariancja:  $\sigma^2 = D^2(X) = E((X-m)^2)$ , odchylenie standardowe:  $\sigma = \sqrt{\sigma^2} = \sqrt{D^2(X)}$ 

Wariancja dla rozkładu dyskretnego:  $D^2(X) = \sum_{i=1}^n (x_i - m)^2 p_i$ , dla rozkładu ciągłego:  $D^2(X) = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx$ Zmienne niezależne gdy dla dowolnych zdarzeń  $B_1, ..., B_k \in \Sigma$ :  $P(X_1 \in B_1, ..., X_k \in B_k) = P(X_1 \in B_1) \cdot ... \cdot P(X_k \in B_k)$ .

Wartości własności i wariancji:

jeżeli X = const = c, to E(X) = c;  $E(aX) = aE(X) \forall a \in \mathbb{R}$ ; E(X+Y) = E(X) + E(Y);  $D^2(X) = E(X^2) - E(X)^2$ ;  $D^2(aX) = a^2D^2(X) \forall a \in \mathbb{R}$ X = const = c to  $D^2(X) = 0$ ; jeżeli X i Y są niezależnymi zmiennymi losowymi, to  $D^2(X + Y) = D^2(X) + D^2(Y)$ 

### Rozkłady

Rozkład Bernouliego:  $\binom{n}{k} p^k (1-p)^{n-k}$ , m=np,  $\sigma^2=np(1-p)$ 

Jeżeli  $X \sim B(n,p)$  i  $Y \sim B(m,p)$  są dwiema niezależnymi zmiennymi losowymi o rozkładzie dwumianowym, wtedy ich suma X+Y jest zmienną losową o rozkładzie dwumianowym B(n+m,p)

Rozkład Poissona (duża liczba prób bernoullego  $m \geqslant 100 \land$  z małym prawdopodobieństwem  $p \leqslant \frac{1}{10}): f(x) = \frac{e^{-\lambda} \lambda^k}{k!}, \lambda = np, m = \lambda, \sigma = \lambda$  Dla dwóch zmiennych losowych o rozkładzie Poissona z parametrami  $\lambda$  i  $\mu$  suma tych zmiennych losowych ma rozkład Possiona o parametrze  $\lambda + \mu$ Rozkład jednostajny:  $f(x) = \frac{1}{b-a}$  gdy  $x \in [a,b], 0$  gdy  $x \notin [a,b], F(x) = 0$  gdy x < a  $\frac{x-a}{b-a}$  gdy  $x \in [a,b]$  1 gdy  $x > b, m = \frac{a+b}{2}, \sigma = \frac{(b-a)^2}{12}$ 

Rozkład wykładniczy (ciągły odpowiednik rozkładu geometrycznego - czas oczekiwania na pierwszy sukces w nieskończonym ciągu niezależnych prób Bernolulliego):  $f(x) = \lambda e^{-\lambda x}, x \geqslant 0, \ F(x) = 1 - e^{-\lambda x}, \ m = \frac{1}{\lambda}, \ \sigma = \frac{1}{\lambda^2}$  Rozkład geometryczny:  $p(1-p)^{k-1}, \ m = \frac{1}{p}, \ \sigma = \frac{1-p}{p^2}$ 

Rozkład normalny:  $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$  Własności  $\Phi(x)$ :  $\Phi(x) = 1 - \Phi(-x)$ ,  $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$ ,  $\Phi(0) = \frac{1}{2}$ ,  $\Phi_{m,\sigma}(x) = \Phi(\frac{x-m}{\sigma})$ 

Dla X bedącego zmienną losową o rozkładzie normalnym  $N(m,\sigma)$  i Y=aX+b, gdzie  $a\neq 0$  Y ma rozkład normalny  $N(am+b,|a|\sigma)$ 

Dystrybuanta zmiennej losowej:  $F(x) = F_X(x) = P_X((-\infty, x]) = P(X \in (-\infty, x])$ . Pochodna dystrybuanty to funkcja rozkładu: F'(x) = f(x) Dystrybuanta jest niemalejąca,  $\lim_{x\to-\infty} F(x) = 0$ ,  $\lim_{x\to\infty} F(x) = 1$ . Dla rozkładu dyskretnego:  $F(x) = \sum_{i:x_i \leqslant x} p_i$ 

#### Centralne twierdzenie graniczne

Dla  $S_n = X_1 + ... + X_n$ , gdzie  $X_i$  to niezależne zmienne losowe z tym samym rozkładem, nadzieją m i wariancją  $\sigma^2$ ,  $\sigma > 0$ :  $Z_n = \frac{S_n - E(S_n)}{\sqrt{D^2(S_n)}} = \frac{S_n - nm}{\sigma \sqrt{n}} - Z_n \text{ to standaryzacja sumy } S_n, E(Z_n) = 0, D^2(Z_n) = 1 \text{ tw. Lindeberga-Levy'ego: } \forall x \in \mathbb{R} \lim_{n \to \infty} P(Z_n \leqslant x) = \Phi(x)$ 

 $\sqrt{D^2(S_n)} \qquad \forall x \in \mathbb{R} \text{ lim}_{n \to \infty}(F_{S_n}(x) - \Phi_{nm,\sigma\sqrt{n}}(x)) = 0. \text{ Dla średnich: } \forall x \in \mathbb{R} \text{ lim}_{n \to \infty}(F_{\underline{S_n}}(x) - \Phi_{m,\frac{\sigma}{\sqrt{n}}}(x)) = 0.$ 

tw. de Moivre'a-Laplace'a (gdy  $X_i$  to ciąg niezależnych prób Bernoullego z tym samym p):  $\forall x \in \mathbb{R} \ P\left(\frac{S_n - np}{\sqrt{npq}} \leqslant x\right) \to \Phi(x)$ :  $\Phi(\frac{S_n - np}{\sqrt{npq}}), \ q = 1 - p$ 

# Estymacja punktowa

Niech  $X_1,...,X_n$  będzie próbką prostą ze zmiennej losowej X. Estymatorem parametru  $\theta$  rozkładu  $P_{\theta} \in \mathbb{P}$  "odpowiednio bliskiego" rozkładowi  $P_X$  nazywamy zmienną losową  $\hat{\theta} \circ (X_1,...,X_n) = T(X_1,...,X_n)$  gdzie T jest odpowiednio dobraną funkcją, która "rozsądnie" przybliża (estymuje)

wartość  $\theta$ . Przykładami estymatorów są: średnia arytmetyczna z próbki -  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  - estymator wartości oczekiwanej E(X). Mediana z próbki -  $meX_{(\lceil n/2 \rceil)}$  - estymator mediany. Wariancja z próbki -  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$  (jeżeli E(X) = m jest znane), lub  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  (jeżeli E(X) = m nie jest znane) - estymator wariancji  $D^2(X)$ 

Estymator nieobciążony -  $E(\hat{\theta}) = \theta$ . Estymator zgodny -  $P(\omega \in \Omega : \lim_{n \to \infty} \hat{\theta}_n(\omega) = \theta) = 1$  Metoda MLE: dla zmiennych losowych  $X_1, ..., X_n$ ,  $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i)$  - dla zmiennych dyskretnych,  $L(\theta) = \prod_{i=1}^{n} f(x_i)$  - dla zmiennych ciągłych, Żeby wyznaczyć MLE  $(\theta)$  należy wyznaczyć maximum funkcji wiarygodności  $L(\theta)$ . Kwantyle i kwartyle -  $q_p = x(\lceil pn \rceil) = min\{x : F_n(x) \geqslant p\}, Q_k = q_{k/4}(k=1,2,3)$ . Klasyczny współczynnik

asymetrii  $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$ , pozycyjny współczynnik asymetrii -  $\frac{Q_3+Q_1-2me}{IQR}$ . Rozstęp międzykwartylowy -  $IQR=q_{3/4}-q_{1/4}$ .  $mad=me(|x_1-me|,...,|x_n-me|)$ .

#### Przedziały ufności Estymacja Przedziałowa

Dla wartości oczekiwanej w rozkładzie normalnym ze znanym odchyleniem standardowym (na poziomie ufności  $1-\alpha$ ):

 $(\bar{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2}), \bar{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})), (-\infty, \bar{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha)), (\bar{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha), \infty)$ 

Dla wartości oczekiwanej w rozkładzie normalnym z nieznanym odchyleniem standardowym:  $(\bar{X} - \frac{S}{\sqrt{n-1}}F_{t_{n-1}}^{-1}(1 - \frac{\alpha}{2}), \bar{X} + \frac{S}{\sqrt{n-1}}F_{t_{n-1}}^{-1}(1 - \frac{\alpha}{2})), (-\infty, \bar{X} + \frac{S}{\sqrt{n-1}}F_{t_{n-1}}^{-1}(1 - \alpha)), (\bar{X} - \frac{S}{\sqrt{n-1}}F_{t_{n-1}}^{-1}(1 - \alpha), \infty),$  Dla frakcji. Próbka prosta  $X_1, ..., X_n$  pochodzi z rozkładu dwupunktowego B(1, p). W przypadku. Dla próbki dużej (n > 30):  $(\hat{p} - \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2}), \hat{p} + \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})), (0, \hat{p} + \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\Phi^{-1}(1 - \alpha), (\hat{p} - \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\Phi^{-1}(1 - \alpha), 1) \text{ gdzie } \hat{p} = \bar{X}_n = \frac{\#\{i: X_i = 1\}}{n}$  Dla wariancji w rozkładzie normalnym z nieznaną wartością oczekiwaną:  $(\frac{nS^2}{F_{n-1}^{-1}(1-\frac{\alpha}{2})}, \frac{nS^2}{F_{n-1}^{-1}(\frac{\alpha}{2})}), (0, \frac{nS^2}{F_{n-1}^{-1}(1-\alpha)}, \infty)$  Przedziały ufności dla wartości oczekiwanej - uwagi. Jeżeli rodzina rozkładow nie jest znana oraz proka giest duża  $(n \geqslant 30)$ , to konstruując

przedziały ufności dla wartości oczekiwanej m możemy rozważyć zmienną losową  $Z = \frac{\bar{X} - m}{c} \sqrt{n} \approx N(0, 1)$ 

Jeżeli natomiast próbka jest mała (n < 30) oraz pochodzi z rozkładu B(1,p) to konstruując przedział ufności dla p możemy rozważyć zmienną losową  $K = \#\{i : X_i = 1\} \sim B(n, p)$ 

#### Testowanie hipotez statystycznych

```
Próbka X_1, ..., X_n z rozkładu N(m, \sigma), stat. testowe dają zm. los. przy prawdziwości hipotez zerowych.
```

Testowanie hipotez 
$$H_0: m=m_0$$
o wart. oczekiwanej w rozkładzie norm.: gdy  $\sigma$  znana  $z=z(x_1,...,x_n)=\frac{\bar{x}-m_0}{\sigma}\sqrt{n}$  dająca zm. los.  $Z=z(X_1,...,X_n)$  o rozkładzie  $N(0,1)$ ,

gdy  $\sigma$  zimini  $z=z(x_1,...,x_n)=\frac{\sigma}{s}$   $\sqrt{n}$  dająca ziminis:  $Z=z(x_1,...,x_n)$  o rozkładzie tv(0,1), gdy  $\sigma$  nieznana  $t=t(x_1,...,x_n)=\frac{\bar{x}-m_0}{s}\sqrt{n-1}$  dająca ziminis:  $T=t(X_1,...,X_n)$  o rozkładzie t-studenta o n-1 st. swobody. Testowanie hipotez  $H_0: \sigma^2=\sigma_0^2$  o wariancji w rozkładzie norm.:

$$\chi=\chi(x_1,...,x_n)=rac{ns^2}{\sigma_0^2}$$
dająca zm. los.  $\chi=\chi(X_1,...,X_n)$  o rozkładzie  $\chi$  a o  $n-1$  st. swobody.

Testowanie hipotez $\check{H_0}:p=p_0$ o frakcji,

gdy 
$$X_1,...,X_n$$
 z rozkładu  $B(1,p)$ : dla próbki  $n\geqslant 30$  używamy stat. testowej  $z=z(x_1,...,x_n)=\frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})}})$  dająca zm. los.  $Z=z(X_1,...,X_n)$  o rozkładzie  $N(0,1)$ ,

dla małej próbki stat. testowa  $k=k(x_1,...,x_n=\#\{i:x_i=1\}$  dająca zm. los.  $K=k(X_1,...,X_n)$  o rozkładzie  $B(n,p_0)$ .

Test t-Studenta: próbki z rozkł.  $N(m_1, \sigma_1)$  i  $N(m_2, \sigma_2)$ ,  $H_0: m_1 = m_2$ ,

dla znanych 
$$\sigma_1, \sigma_2$$
:  $z = z(x_1, ..., x_{n_1}, y_1, ..., y_{n_2}) = \frac{\hat{x} - \hat{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ ) dająca zm. los.  $Z$  o rozkł.  $N(0, 1)$ 

Test t-Studenta: probki z rozki.  $N(m_1, \sigma_1)$  i  $N(m_2, \sigma_2)$ ,  $m_0 \cdot m_1 - m_2$ , dla znanych  $\sigma_1, \sigma_2$ :  $z = z(x_1, ..., x_{n_1}, y_1, ..., y_{n_2}) = \frac{\hat{x} - \hat{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$  dająca zm. los. Z o rozkł. N(0, 1), dla nieznanych  $\sigma_1, \sigma_2$ :  $t = t(x_1, ..., x_{n_1}, y_1, ..., y_{n_2}) = \frac{\hat{x} - \hat{y}}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} \cdot \frac{n_1 + n_2}{n_1 n_2}}}$  dająca zm. los. T o rozkładzie t-studenta o  $n_1 + n_2 - 2$  st. swobody.

 $\text{Test}\chi^2$ zgodnosci: dla rozkładów dyskretnych:  $H_0: P_X = P$   $P(y_1) = \pi_1 > 0, \, ..., \, P(y_k) = \pi_k > 0, \, \pi_1 +, \, ... \, , + \pi_k = 1, \, n_i \text{ - liczba wystąpień } y_i \text{ w ciągu } x_1, ..., x_n$ 

$$\chi = \chi((x_1, y_1), ..., (x_n, y_n)) = \sum_{(s,t) \in S \times T} \frac{(n_{s,t} - \frac{n_s n_t}{n})^2}{\frac{n_s n_t}{n}}, \text{ gdzie } S \times T \text{ jest nośnikiem próbki danych,}$$

$$P(y_1) = \pi_1 > 0, \dots, P(y_k) = \pi_k > 0, \pi_1 + \dots + \pi_k = 1, n_i - \text{liczba wystąpien } y_i \text{ w ciągu } x1, \dots, x_n$$

$$\chi = \chi(x_1, \dots, x_n) = \sum_{i=1}^k \frac{(n_i - n\pi_i)^2}{n\pi_i}, \text{ zbiór krytyczny K to } [l, \text{inf}), \text{ gdzie l to kwantyl rzędu } 1 - \alpha \text{ rozkładu } \chi^2 \text{ o } k - 1 \text{ stopniach swobody.}$$
Test niezależności rozkładów: dla zmiennych losowych:  $X \sim P_X$  i  $Y \sim P_Y$ 

$$\chi = \chi((x_1, y_1), \dots, (x_n, y_n)) = \sum_{(s,t) \in S \times T} \frac{(n_{s,t} - \frac{n_s n_t}{n})^2}{\frac{n_s n_t}{n}}, \text{ gdzie } S \times T \text{ jest nośnikiem próbki danych,}$$

$$n_{s,t} = \#\{i : (s,t) = (x_i, y_i)\}, n_s = \#\{i : s = x_i\}, n_t = \#\{i : t = y_i\} \ (n = \sum_{(s,t)} n_{s,t}) \text{ Można wtedy wykazać, że } \chi \approx \chi^2_{(\#S-1)(\#T-1)}$$

# Metoda bootstrap

Dla małej próbki (o wielkości n) i nieznanym rozkładzie losujemy z niej ze zwracaniem kolejno B próbek o wielkości n.

Estymator bootstrapowy parametru ze znanym estymatorem  $g(x_1,...,x_n)$  to:  $\hat{g}=\hat{g}(x_1,...,x_n)=\frac{1}{B}\sum_{i=1}^Bg(x_1^i,...,x_n^i)$ , gdzie  $x_1^i,...,x_n^i$  to próbka wylosowana za i-tym razem.

Metoda percentylowa wyznaczania przedziałów ufności parametru  $\theta$ : losujemy 1000 próbek bootstrapowych, dla każdej obliczamy estymator  $\theta$ . Kwantyle odpowiednich rzędów z ciągu estymatorów dla próbek są końcami przedziału ufności.

#### Wektor losowy

Wektor losowy: funkcja  $X: \Omega \to \mathbb{R}^n \ (Y: \Omega \to \mathbb{R}^m)$  na przestrzeni  $(\Omega, \Sigma, P)$ , rozkład wektora losowego  $X: P_X(B) = P(X^{-1}(B) \ \text{dla} \ B \subset \mathbb{R}^n$ . Dla  $A_1 \subset \mathbb{R}^n A_2 \subset \mathbb{R}^m$ :  $P_X(A_1) = P_{(X,Y)}(A_1 \times \mathbb{R}^m)$  i  $P_Y(A_2) = P_{(X,Y)}(\mathbb{R}^n \times A_2)$  są rozkładami brzegowymi, a  $P_{(X,Y)}$  to rozkład łączny. Niezależność wektorów losowych o rozkładach ciągłych  $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$ 

dla  $x \in \mathbb{R}^n, y \in \mathbb{R}^m, P_X(x) > 0, P_Y(y) > 0, f_X(x) > 0, f_Y(y) > 0$ 

Rozkłady warunkowe wektora losowego (dyskretny):  $P_{X|Y=y}(B) = P(X \in B|Y=y) = \frac{P(X \in B, Y=y)}{P(Y=y)}$  dla  $B \subset \mathbb{R}^n$ 

Rozkłady warunkowe wektora losowego (ciągłego):  $f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_{X}(x)}$  dla  $y \in \mathbb{R}^m$ 

Warunkowa wartość oczekiwana: E(X|Y=y)

#### 10 Regresja Liniowa

Model regresji liniowej:  $Y_i = \alpha + \beta x_i + U_i$  dla i = 1,..,n,

Wyznaczenie estymatorów  $\alpha$  i  $\beta$  MNK: wyznaczamy arg min  $S(\alpha, \beta)$  dla $S(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$ ,

otrzymujemy  $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$   $\hat{\beta} = \frac{n\Sigma_{i=1}^n x_i y_i - (\Sigma_{i=1}^n x_i)(\Sigma_{i=1}^n y_i)}{n\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)(\Sigma_{i=1}^n y_i)}$   $\hat{\alpha}$  i  $\hat{\beta}$  są nieobciążone. Wyznaczenie estymatorów metodą największej wiarygodności dla błędów normalnych:

Zał:  $U_i \sim N(0, \sigma)$ , czyli  $Y \sim N(\alpha + \beta x_i, \sigma)$ 

$$L(\alpha,\beta,\sigma^2) = f_1(y_1) \cdots f_n(y_n), \text{ gdzie } f_i(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\alpha-\beta x_i)^2}{(2\sigma^2)}} \text{ dostajemy te same estymatory jak w MNK oraz } \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \text{ a także } E(\hat{\sigma^2}) = \frac{n-2}{n} \sigma^2$$

# Analiza wariancji (ANOVA)

Rozkład F(-Snedecora): Niech X i Y będą niezależnymi zmiennymi losowymi o rozkładach  $\chi_p^2$  i  $\chi_q^2$ 

(k-1,n-k) stopniach swobody.  $K=[a,\infty),$  a - kwantyl rzędu  $1-\alpha$  rozkładu F-Snedecora.

Zatem  $F = \frac{X/p}{Y/q}$  posiada rozkład F-Snedecora o (p,q) stopniach swobody, jeżeli T jest zmienną losową o rozkładzie  $t_q$ , to  $T^2 \sim F_{1,q}$ ,  $E(F) = \frac{q}{q-2}$ oraz  $D^2(F) = \frac{2q^2(p+q-2)}{p(q-2)^2(q-4)}$  dla q>4. Jednoczynnikowa analiza wariancji: Mając k niezależnych próbek prostych:  $X_{11},...,X_{1n_1},X_{21},...,X_{2n_2},...,X_{k1},...,X_{kn_k}$  które pochodzą z  $N(m_1,\sigma),...,N(m_k,\sigma)$  testujemy hipotezę:  $H_0: m_1=m_2=...=m_k$  wobec  $H_1:$  nie wszystkie wartości  $m_i$  są sobie równe. Do weryfikacji  $H_0$  służy  $f=\frac{MSTR}{MSE},$   $MSTR=\frac{1}{k-1}\sum_{i=1}^k n_i(\bar{x_i}-\bar{x})^2,$   $MSE=\frac{1}{n-k}\sum_{i=1}^k n_is_i^2.$   $n=\sum_{i=1}^k n_i,\bar{x_i}$  jest średnią arytmetyczną z i-tej próbki,  $s_i^2$  jest wariancją z i-tej próbki,  $\bar{x}$  jest średnią arytmetyczną ze wszystkich obserwacji, która daje  $F = F(X_{11},..,X_{kn_k})$  o rozkładzie F-Snedecora o