# Wykład 12 (26.05.08): Testy dla dwóch prób w rodzinie rozkładów normalnych

**Przykład** Rozważamy dane wygenerowane losowo; ( podobne do danych z przykładu 7.2 z książki A. Łomnickiego)

$n_1 = 9$ poletek w dąbrowie, $n_2 = 10$ poletek w borze
Liczby pająków (na poletkach) w dąbrowie : 45, 57, 30, 61, 52, 64, 50, 60, 48;

Liczby pająków (na poletkach) w borze: 46, 32, 39, 34, 46, 31, 37, 26, 52, 21.

Chcielibyśmy zweryfikować hipotezę o równości liczby pająków w borze i dąbrowie.

# Formalizacja problemu

Zakładamy,że $x_1, x_2, \ldots, x_{n_1}$, liczby pająków na poletkach w dąbrowie, są realizacjami próby prostej $X_1, X_2, \ldots, X_{n_1}$,

$$X_1, X_2, \ldots, X_{n_1}, \quad X_i \sim N(\mu_1, \sigma_1), \quad i = 1, 2, \ldots, n_1$$

oraz że $y_1, y_2, \ldots, y_{n_2}$ są realizacjami próby prostej

$$Y_1, Y_2, \ldots, Y_{n_2}, \quad Y_i \sim N(\mu_2, \sigma_2), \quad i = 1, 2, \ldots, n_2.$$

W naszym przykładzie :

$$x_1 = 45, \ldots, y_1 = 46, \ldots.$$

Chcemy zweryfikować hipotezę

$$H_0 : \mu_1 = \mu_2 \textit{ przeciwko hipotezie } H_1 : \mu_1 \neq \mu_2$$

## Weryfikacja hipotezy o równości wariancji

Przy założeniu, że: $\sigma_1 = \sigma_2$ (warunek "jednorodności wariancji" jest spełniony) moglibyśmy zastosować test Studenta dla jednorodnych wariancji dla prównania dwóch średnich. Ma on wiele zalet (z punktu widzenia teorii testowania hipotez). Dlatego pierwszym krokiem powinno być sprawdzenie hipotezy o równości wariancji:

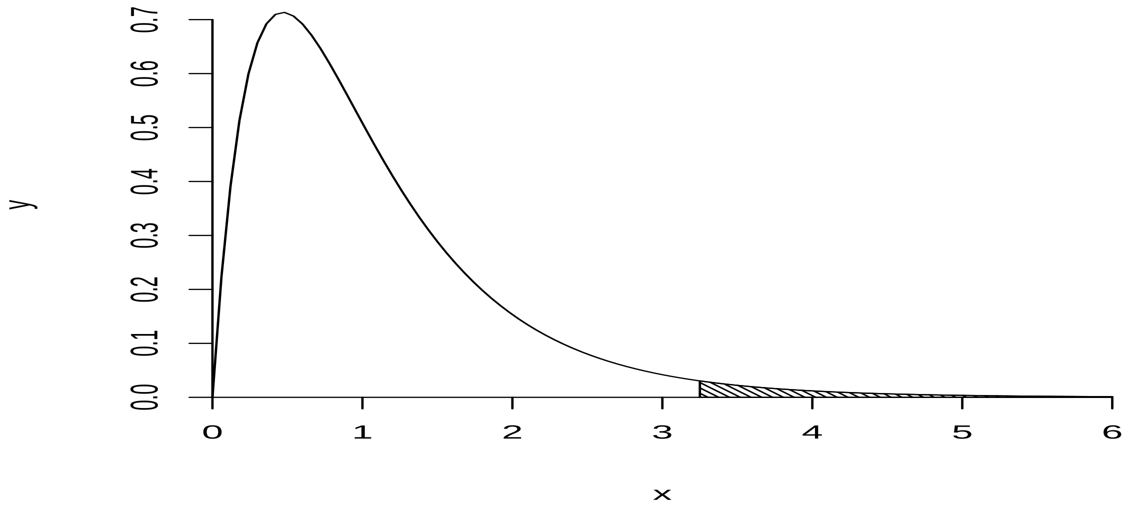
$$H_0^j : \sigma_1^2 = \sigma_2^2 \text{ przeciwko } H_1^j : \sigma_1^2 \neq \sigma_2^2.$$

Odpowiedną statystyką testową dle weryfikacji $H_0^j$ przeciwko $H_1^j$ jest

$$F = \frac{\mathcal{S}_1^2}{\mathcal{S}_2^2}$$

gdzie $\mathcal{S}_1^2$ jest wariancją próbkową dla pierwszej próby, $\mathcal{S}_2^2$ jest wariancją próbkową dla drugiej próby. Można pokazać, że $F$ ma rozkład $F_{n_1-1,n_2-1}$ (rozkład F-Snedecora z $n_1 - 1$ i $n_2 - 1$ stopniami swobody).

# Rozkład F-Snedecora




Rysunek 1: Wykres gęstości rozkładu rozkładu $F_{4,30}$ ; obszar zakreskowany odpowiada wartościom większym niż $F_{0,975;4;30} = 3{,}25$.

Funkcja gęstości rozkładu $F_{n_1,n_2}$ patrz Koronacki i Mielniczuk (2001, str. 208).

# Hipoteza o równości wariancji— obszar krytyczny

Obszar krytyczny, dla poziomu istotności $\alpha$, ma postać

$$[0, F_{\alpha/2;n_1,n_2}] \cup [F_{1-\alpha/2,n_1,n_2}, \infty).$$

Weryfikacja

$$H_0^j : \sigma_1^2 = \sigma_2^2 \text{ przeciwko } H_1^j : \sigma_1^2 \neq \sigma_2^2.$$

może być również oparta na statystyce

$$F^{'} = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)};$$

obszar krytyczny (odpowiadający $F^{'}$) ma postać:

$$[F_{1-\alpha/2,n_1,n_2}, \infty)$$

**Przykład — obliczenia**

$$s_1^2 = 108{,}3611; \quad s_2^2 = 92{,}7111;$$

widzimy, że $s_1^2 \geqslant s_2^2$;

$f^{'}$, realizacja statystyki testowej $F^{'}$, jest równa

$$f^{'} = \frac{s_1^2}{s_2^2} = 1{,}168804.$$

Wartość kwantyla rzędu $0{,}975$ rozkładu $f_{8,9}$ (oznaczmy ją przez $f_{0,975,8,9}$) jest równa

$$f_{0,975,8,9} = 4{,}101956,$$

wartość statystyki testowej nie należy do obszaru krytycznego $[4{,}101956, \infty)$, więc nie ma podstaw do odrzucenia hipotezy $H_0^j$.

## Testowanie hipotezy o równości średnich — przypadek jednorodnych wariancji

Jeśli $X_1, X_2, \ldots, X_{n_1}$ jest próbą prostą, $X_i \sim N(\mu_1, \sigma)$, a $Y_1, Y_2, \ldots, Y_{n_2}$ jest również próbą prostą , $Y_i \sim N(\mu_2, \sigma)$, (zauważ, że wariancje obu rozkładów, z których pochodzą próby, są równe!), wtedy można do weryfikacji hipotezy

$$H_0 : \mu_1 = \mu_2 \quad \text{przeciwko} \quad H_1 : \mu_1 \neq \mu_2$$

można zastosować "klasyczny" test t-Studenta dla dwóch prób.

Statystyka testowa ma postać:

$$T = \frac{|\bar{X} - \bar{Y}|}{S_{X,Y}},$$

gdzie

$$\mathcal{S}_{X,Y} = \sqrt{\frac{(n_1-1)\mathcal{S}_1^2 + (n_2-1)\mathcal{S}_2^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}},$$

$\mathcal{S}_1^2$ oznacza wariancję zmiennych $X_i$, $i = 1, \ldots, n_1$, $\mathcal{S}_2^2$ oznacza wariancję zmiennych $Y_i$, $i = 1, \ldots, n_2$,

a obszar krytyczny, dla poziomu istotności $\alpha \in (0,1)$ :

$$[t_{1-\alpha/2, n_1+n_2-2}, \infty).$$

# Przykład — obliczenia

Przyjmujemy $\alpha = 0{,}05$ (poziom istotności testowania hipotezy);

$$
\begin{aligned}
n_1 &= 9, \quad n_2 = 10 \\
\bar{x} &= 51{,}88889 \quad \bar{y} = 36{,}4 \\
s_1^2 &= 108{,}3611, \quad s_2^2 = 92{,}71111 \\
s_{x,y}^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \times \frac{n_1+n_2}{n_1 n_2} = \\
&= \frac{8 \times 108{,}3611 + 9 \times 92{,}71111}{9+10-2} \times \frac{9+10}{9 \times 10} = \\
&= 21{,}12712 \\
s_{x,y} &= \sqrt{21{,}12712} = 4{,}596424 \\
t &= \frac{|\bar{x}-\bar{y}|}{s_{x,y}} = \frac{51{,}88889 - 36{,}4}{4{,}596424} \approx 3{,}369769.
\end{aligned}
$$

Odczytujemy z tablic rozkładu t-Studenta (lub obliczamy korzystając z R-owskiego polecenia **qt**)

$$t_{0,975;17} = 2{,}109816.$$

Wartość statystyki testowej należy do obszaru krytycznego, a zatem należy odrzucić hipotezę o równości liczebności pająków na jednostkę powierzchni w borze i dąbrowie.

# Obliczenia w środowisku R

Powyższe obliczenia można wykonać przy pomocy pakietu R w następujący sposób:

```
> x
[1] 45 57 30 61 52 64 50 60 48
> y
 [1] 46 32 39 34 46 31 37 26 52 21
> t.test(x,y,var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = 3.3698, df = 17, p-value = 0.003638
alternative hypothesis:
```

```
true difference in means is not equal to 0
95 percent confidence interval:
  5.791281 25.186497
sample estimates:
mean of x mean of y
 51.88889  36.40000
```

## Wersje testu t dla przypadku niejednorodnych wariancji

Jeśli nie można przyjąć założenia o jednorodności wariancji, wtedy należy użyć wersji testu t dla przypadku niejednorodnych wariancji takich jak:
- test oparty na statystyce Cochrana-Coxa (por. Łomnicki, Rozdz. 7)
- test oparty na aproksymacji Welcha (dostępny w pakiecie R)

Należy pamiętać, że korzystanie z powyższych testów wymaga założenia normalności porównywanych populacji.

# Obliczenia przy braku założenia dotyczącego równości wariancji

```
> t.test(x,y,var.equal=FALSE)

        Welch Two Sample t-test

data:  x and y
t = 3.3552, df = 16.413, p-value = 0.003907
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
  5.722523 25.255255
sample estimates:
mean of x mean of y
 51.88889  36.40000
```

## Obliczenia przy braku założenia dotyczącego równości wariancji — komentarz

Wartość statystyki testowej $t$- obliczana jest trochę inaczej niż w „przypadku równych wariancji”;
liczba stopni swobody— nie jest liczbą całkowitą!

## Polecana literatura

J. Koronacki i J. Mielniczuk, Statystyka dla studentów kierunków technicznych i przyrodniczych, WNT 2001, str. 229-237.

A. Łomnicki, Wprowadzenie do statystyki dla przyrodników, Wyd. 3, Rozdział 7.