



## Ćwiczenie: Testy nieparametryczne dla dwóch i wielu próbek

### Zadanie 1

1. Przekopij skoroszyt Excela o nazwie **nieparam.xls** do własnego folderu.  
*Skoroszyt ten zawiera dane dotyczące wybranych „wskaźników jakości” ścieków trafiających do oczyszczalni oraz produktu powstającego po oczyszczeniu. Dane te pochodzą z różnych miesięcy, te zaś zostały podzielone na pory roku: zima, wiosna, lato.*
2. Traktując eksperyment jako doświadczenie dwugrupowe (wyłącz z porównania porę roku „lato”) porównaj wskaźniki ścieków otrzymane zimą i wiosną. W tym celu zastosuj test U Manna-Whitneya.
3. Wykonaj statystyczną analizę biorąc już pod uwagę wszystkie pory roku. W razie potrzeby wykonaj analizę wariancji lub test Kruskala-Wallisa (nieparametryczny odpowiednik analizy wariancji).

### Poniżej znajduje się przykładowe rozwiązanie dotyczące ChZT0 wykonane w programie SAS.

#### 1. Badamy czy zmienna posiada rozkład zgodny z normalnym

Sprawdzamy czy zmienna posiada rozkład zgodny z normalnym, czyli weryfikujemy  $H_0$ .

$H_0$ : Zmienna ChZT posiada rozkład zgodny z normalnym

$H_1$ : Zmienna ChZT nie posiada rozkładu zgodnego z normalnym

```
                The UNIVARIATE Procedure
Variable:   ChZT0

                Moments

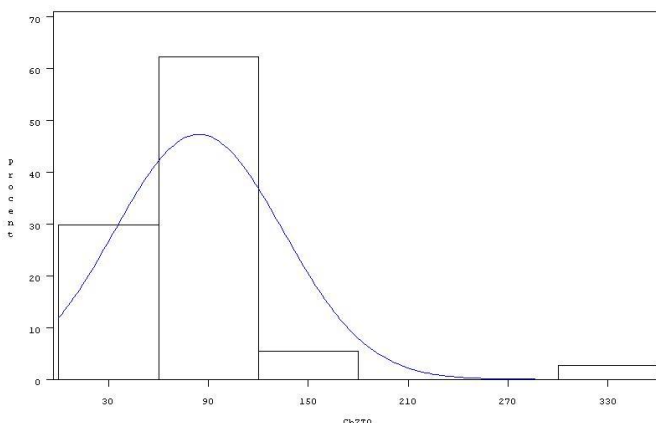
      N              37      Sum Weights              37
      Mean           84.2702703      Sum Observations      3118
      Std Deviation  50.6379078      Variance            2564.1977
      Skewness       3.65602344      Kurtosis           17.673983
      Uncorrected SS 355065.82      Corrected SS       92311.1173
      Coeff Variation 60.0898841      Std Error Mean     8.32482076
```

```
                Testy normalności

      Test              --Statystyka--      ----p Value-----

      Shapiro-Wilk      W      0.651028      Pr < W      <0.0001
      Kolmogorov-Smirnov D      0.181491      Pr > D      <0.0100
      Cramer-von Mises W-Sq  0.416672      Pr > W-Sq   <0.0050
      Anderson-Darling  A-Sq  2.738474      Pr > A-Sq   <0.0050
```

Obliczone prawdopodobieństwo wynikające z przeprowadzonego testu Shapiro-Wilk ( $p < 0,0001$ ) upoważnia nas do odrzucenia  $H_0$ . Stwierdzamy zatem, że zmienna ChZT0 nie posiada rozkładu zgodnego z normalnym. Upewnia nas o tym również poniższy wykres typu histogram.





## 2. Obliczamy miary położenia klasyczne i pozycyjne

	ChZT0		
	pora		
	lato	wiosna	zima
N	16	13	8
Mean	73.89	110.56	62.30
Median	69.70	94.80	61.05
Q1	50.80	72.70	47.05
Q3	93.20	116.50	67.20
Std	28.78	72.27	19.40

## 3. Porównujemy pory roku

a)

The screenshot shows the Minitab software interface. The 'Zadania' (Tasks) menu is open, and the 'ANOVA' option is selected. The 'ANOVA' submenu is visible, showing options like 'Test t...', 'ANOVA jednoczynnikowa...', and 'Nieparametryczna ANOVA jednoczynnikowa...'. Below the menu, a data table is visible with columns for 'miesiac' (month) and 'pora' (season). The data rows are as follows:

miesiac	pora
6 luty	zima
7 luty	zima
8 luty	zima
9 marzec	wiosna
10 marzec	wiosna
11 marzec	wiosna
12 marzec	wiosna
13 kwiecień	wiosna
14 kwiecień	wiosna

b)

The screenshot shows the 'Dane' (Data) dialog box in Minitab. The 'Źródło danych' (Data source) is set to 'Local:WORK.NIEPARAM'. The 'Filtr zadań' (Task filter) is set to 'Brak' (None). The 'Zmienne przypisywane' (Assigned variables) section shows the following variables: LP, miesiac, pora, ZawOg0, and ChZT0. The 'Role zadania' (Task roles) section shows the following roles: Zmienne zależne (Dependent variables), ChZT0, Zmienna niezależna (Limit: 1) (Independent variable), pora, Grupuj analizowane wg (Group analysis by), and Liczebność (Limit: 1) (Count).



c)

## Wyniki:

### Test sum rang Wilcozona jest odpowiednikiem testu U Mann-Whitney'a

Porównujemy poziom ocenianych wskaźników ścieków między zimą i wiosną. Weryfikujemy zatem hipotezę zerową zakładającą, iż rozkład ChZT stwierdzony zimą i wiosną jest taki sam:

$H_0: F(x) = G(x)$ ;  $H_1: F(x) \neq G(x)$

$F(x)$ ,  $G(x)$  – dystrybuanta ChZT zimą i wiosną

Wcześniej z importowanego pliku **nieparam** należy utworzyć podzbiór zawierający dane dotyczące zimy i wiosny. Powyższy efekt uzyskamy na drodze **Budowy zapytań**.

```
The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) dla zmiennej ChZT0
klasyfikowanej wg zmiennej pora

      pora      N      Suma      Oczekiwane      Odch. std.      Średnia
      pora      ocn      ocen      poniżej H0      poniżej H0      ocena
-----
zima          8          48.0          88.0          13.803726          6.000000
wiosna       13         183.0         143.0          13.803726          14.076923
```

Dla więzów użyto ocen przeciętnych.

#### Wilcoxon Two-Sample Test

Statystyka 48.0000

#### Normal Approximation

Z -2.8615  
One-Sided Pr < Z 0.0021  
Two-Sided Pr > |Z| **0.0042**

#### t Approximation

One-Sided Pr < Z 0.0048  
Two-Sided Pr > |Z| 0.0096

Z includes a continuity correction of 0.5.

#### Kruskal-Wallis Test

Chi-kwadrat 8.3971  
Stopień swobody 1  
Pr > Chi-kwadrat 0.0038

Otrzymane prawdopodobieństwo pozwala odrzucić  $H_0$ . Stwierdzamy zatem, że poziom ChZT0 jest statystycznie różny zimą i wiosną.



### Test Kruskal-Wallis

Tym razem bierzemy pod uwagę wszystkie pory roku.

Sprawdzamy czy pora roku wpływa statystycznie na uzyskane pomiary. Weryfikujemy hipotezę zerową zakładającą, iż rozkład ChZT w k populacjach jest taki sam:

$$H_0: F_1(x) = F_2(x) = \dots = F_k(x)$$

$$H_1: F_1(x) \neq F_2(x) \neq \dots \neq F_k(x)$$

$F_1(x)$ ,  $F_2(x)$ ,  $F_k(x)$  – dystrybuanty rozpatrywanych populacji.

The NPAR1WAY Procedure  
Wilcoxon Scores (Rank Sums) dla zmiennej ChZT0  
klasyfikowanej wg zmiennej pora

pora	N	Suma ocen	Oczekiwane poniżej H0	Odch. std. poniżej H0	Średnia ocena
zima	8	99.0	152.0	27.101522	12.375000
wiosna	13	333.0	247.0	31.428741	25.615385
lato	16	271.0	304.0	32.615146	16.937500

Dla więzów użyto ocen przeciętnych.

#### Kruskal-Wallis Test

Chi-kwadrat 8.4354  
Stopień swobody 2  
Pr > Chi-kwadrat 0.0147

Wartość testu Kruskal-Wallis wynosi 8,4354. Obliczone prawdopodobieństwo ( $p < 0,0147$ ) umożliwia odrzucenie  $H_0$ . Wyniki analizy pozwalają tym samym stwierdzić, że pora roku wpływa statystycznie istotnie na poziom badanego wskaźnika.

### Zadanie 2.

Sprawdź czy noszenie przez 2 tygodnie szkieł kontaktowych wpłynęło na zmianę grubości rogówki oka ludzkiego? (oczy.xls)