

30.05.2018 Gliwice



# Metody Statystyczne

## Sprawozdanie z projektu 18

AEI, Informatyka sem. 4

grupa 6, sekcja 5

Martyna Drabińska

Bartłomiej Krasoń

Michał Miciak

## Treść polecenia

### Projekt 18

Pewien element produkowany jest w nowej i starej hali pewnego zakładu. W ramach badania wydajności pracy (w sztukach na godzinę) przy produkcji tego elementu wylosowano w każdej hali grupę pracowników i wyznaczono ich wydajność pracy. Otrzymano następujące wyniki:

W starej hali zaobserwowano następujące wydajności pracy:

30,22; 34,44; 34,51; 21; 36,59; 32,94; 33,97; 35,7; 38,8; 41,1; 39,87; 35,11; 38,41; 33; 43,31; 32,79; 33,13; 32,92; 37,73; 28,87; 34,73; 39,03; 33,07; 28,98; 29,65; 41,29; 35,05; 26,21; 44,68; 40,78; 39,47; 40,66; 36,62; 33,51; 32,58.

Wydajności pracy w nowej hali były następujące:

42,95; 38,32; 34,9; 25,74; 34,43; 16,58; 36,27; 24,29; 36,81; 48,74; 39,6; 43,79; 32,89; 38,06; 39,8; 33,42; 43,78; 36,4; 35,9; 34,22; 28,59; 33,28; 36,2; 21,56; 35,13; 42,08; 42,85; 38,6; 38,85; 43,23; 38,92; 36,37; 43,36; 44,44; 31,26; 33,71; 41,1; 30,67; 38,88; 34,38; 52,24; 45,25; 51,97; 28,72; 44,74; 24,68; 35,07; 35,43.

## Zadania

1. Dokonać analizy wydajności pracy przy produkcji elementu, wyznaczając miary przeciętne, zróżnicowania, asymetrii i koncentracji. Opracować histogramy rozkładów empirycznych.

Miary wyznaczyć dwoma sposobami:

- a.) na podstawie szeregu szczegółowego
- b.) na podstawie szeregu rozdzielczego

Do wyznaczenia miar szeregów szczegółowych skorzystaliśmy z następujących wbudowanych funkcji\* języka R lub wzorów:

MIARA	SPOSÓB WYZNACZENIA
MIARY PRZECIĘTNE	
Średnia arytmetyczna	<b>mean</b>
Mediana	<b>median</b>
Kwanty Q1	<b>quantile</b>
Kwanty Q3	<b>quantile</b>
MIARY ZRÓŻNICOWANIA	
Wariancja nieobciążona	<b>var</b>
Wariancja obciążona	<b>var* (n - 1)/n</b>
Odchylenie standardowe nieobciążone	<b>sd</b>
Odchylenie standardowe obciążone	<b>sd* (n - 1)/n</b>
Odchylenie przeciętne od średniej	$d_1 = \frac{1}{n} \sum_{i=1}^n  x_i - \bar{x} $
Odchylenie przeciętne od mediany	$d_2 = \frac{1}{n} \sum_{i=1}^n  x_i - Me $

Odchylenie ćwiartkowe	$Q = \frac{1}{2} (Q_3 - Q_1)$
Współczynnik zmienności	$V = \frac{s}{\bar{x}} \cdot 100\%$
MIARY ASYMETRII	
Współczynnik skośności	$A = \frac{M_3}{s^3}$ , gdzie $M_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$
MIARY KONCENTRACJI	
Kurtoza	$K = \frac{M_4}{s^4}$ , gdzie $M_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$
Exces	$q = K - 3$

\* - nazwy funkcji pogróbione

Do wyznaczenia miar szeregów rozdzielczych skorzystaliśmy z następujących wzorów:

Ogólne oznaczenia:

$x_i^*$  - środek  $i$ -tego przedziału

$n_i$  - liczebność  $i$ -tego przedziału

MIARA	SPOSÓB WYZNACZENIA
MIARY PRZECIĘTNE	
Średnia arytmetyczna	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i^* n_i$
Moda	$Mo = x_m + \frac{(n_m - n_{m-1}) h}{(n_m - n_{m-1}) + (n_m - n_{m+1})}$ <p> <math>x_m</math> - lewy koniec przedziału z modą  <math>h</math> - długość przedziału z modą  <math>n_m</math> - liczebność przedziału z modą  <math>n_{m-1}</math> - liczebność przedziału poprzedzającego przedział z modą  <math>n_{m+1}</math> - liczebność przedziału następującego po przedziale z modą </p>
Mediana	$Me = x_{Me} + \frac{h}{n_{Me}} \left( \frac{n}{2} - \sum_{i=1}^{k-1} n_i \right)$ <p> <math>x_{Me}</math> - lewy koniec przedziału z medianą  <math>h</math> - długość przedziału z medianą  <math>n_{Me}</math> - liczebność przedziału z medianą  <math>k</math> - numer przedziału zawierającego medianę </p>

Kwanty Q1	$Q_1 = x_{Q_1} + \frac{h}{n_{Q_1}} \left( \frac{n}{4} - \sum_{i=1}^{k-1} n_i \right)$ <p> <math>x_{Q_1}</math> - lewy koniec przedziału zawierającego <math>Q_1</math>  <math>h</math> - długość przedziału zawierającego <math>Q_1</math>  <math>n_{Q_1}</math> - liczebność przedziału zawierającego <math>Q_1</math>  <math>k</math> - numer przedziału zawierającego <math>Q_1</math> </p>
Kwanty Q3	$Q_3 = x_{Q_3} + \frac{h}{n_{Q_3}} \left( \frac{3n}{4} - \sum_{i=1}^{k-1} n_i \right)$ <p> <math>x_{Q_3}</math> - lewy koniec przedziału zawierającego <math>Q_3</math>  <math>h</math> - długość przedziału zawierającego <math>Q_3</math>  <math>n_{Q_3}</math> - liczebność przedziału zawierającego <math>Q_3</math>  <math>k</math> - numer przedziału zawierającego <math>Q_3</math> </p>
MIARY ZRÓŻNICOWANIA	
Wariancja nieobciążona	$s^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i^* - \bar{x})^2 n_i$
Wariancja obciążona	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x})^2 n_i$
Odchylenie standardowe nieobciążone	$s^* = \sqrt{s^{*2}}$
Odchylenie standardowe obciążone	$s = \sqrt{s^2}$
Odchylenie przeciętne od średniej	$d_1 = \frac{1}{n} \sum_{i=1}^n  x_i^* - \bar{x}  n_i$
Odchylenie przeciętne od mediany	$d_2 = \frac{1}{n} \sum_{i=1}^n  x_i^* - Me  n_i$
Odchylenie ćwiartkowe	$Q = \frac{1}{2} (Q_3 - Q_1)$
Współczynnik zmienności	$V = \frac{s}{\bar{x}} \cdot 100\%$
MIARY ASYMETRII	
Wskaźnik asymetrii	$W_s = \bar{x} - Mo$
Współczynnik skośności	$A = \frac{M_3}{s^3}, \text{ gdzie } M_3 = \frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x})^3 n_i$
MIARY KONCENTRACJI	
Kurtoza	$K = \frac{M_4}{s^4}, \text{ gdzie } M_4 = \frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x})^4 n_i$
Exces	$q = K - 3$

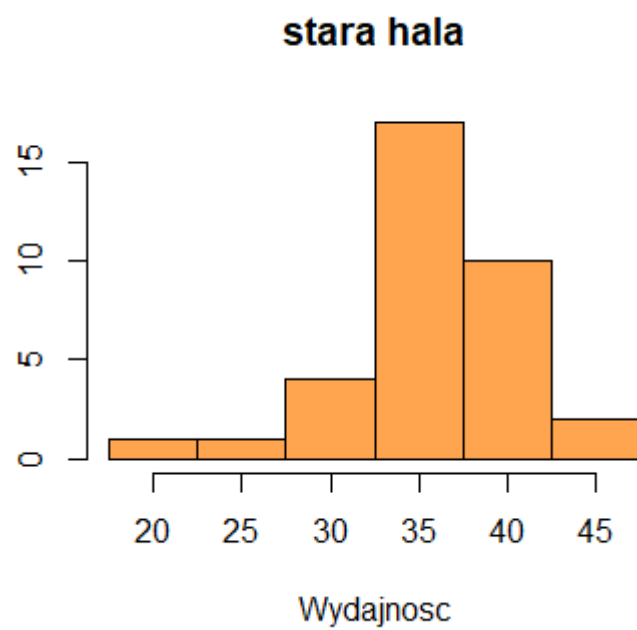
Obliczone miary danych starej hali:

	szczegółowy	rozdzielczy
Średnia arytmetyczna	35.16343	35.71429
Moda		35.75
Mediana	34.73	37.79412
Kwanty Q1	32.93	34.26471
Kwanty Q3	38.915	44
Wariancja nieobciążona	24.59267	24.4898
Wariancja obciążona	23.89002	25.21008
Odchylenie standardowe nieobciążone	4.9591	4.948717
Odchylenie standardowe obciążone	4.817411	5.020964
Odchylenie przeciętne od średniej	3.805061	3.510204
Odchylenie przeciętne od mediany	3.783143	4.163866
Odchylenie ćwiartkowe	2.9925	4.867647
Współczynnik zmienności	13.7%	13.85%
Wskaźnik asymetrii		-0.03571429
Współczynnik skośności	-0.482586	-0.8191157
Kurtoza	3.712131	4.605208
Exces	0.7121305	1.605208

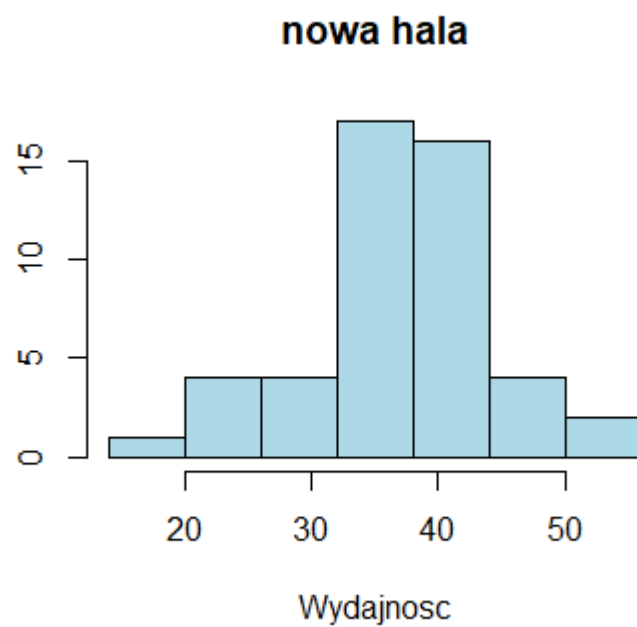
Obliczone miary danych nowej hali:

	szczegółowy	rozdzielczy
Średnia arytmetyczna	36.84271	36.875
Moda		37.57143
Mediana	36.385	37.29412
Kwanty Q1	33.6375	33.05882
Kwanty Q3	42.2725	41.75
Wariancja nieobciążona	53.00418	55.73438
Wariancja obciążona	51.89992	56.92021
Odchylenie standardowe nieobciążone	7.280397	7.465546
Odchylenie standardowe obciążone	7.128722	7.544549
Odchylenie przeciętne od średniej	5.458767	5.78125
Odchylenie przeciętne od mediany	5.438958	5.816176
Odchylenie ćwiartkowe	4.3175	4.345588
Współczynnik zmienności	19.34%	20.24%
Wskaźnik asymetrii		-0.6964286
Współczynnik skośności	-0.374538	-0.3475174
Kurtoza	3.631411	3.365228
Exces	0.6314107	0.3652277

Histogram rozkładu empirycznego dla danych z starej hali:



Histogram rozkładu empirycznego dla danych z nowej hali:



2. Sprawdzić, czy wydajności pracy przy produkcji elementu mają rozkład normalny (test zgodności Kołmogorowa-Lillieforsa, współczynnik ufności 0,95).

$H_0$ : Wydajności pracy przy produkcji elementu mają rozkład normalny.

$H_1$ : Wydajności pracy przy produkcji elementu nie mają rozkładu normalnego.

Wyznaczenie statystyki testowej:

$$d_n = \max(d_n^-, d_n^+)$$

$$\text{gdzie: } d_n^+ = \max_{1 \leq i \leq n} \left| \frac{i}{n} - F_0(x) \right| \quad d_n^- = \max_{1 \leq i \leq n} \left| F_0(x) - \frac{i-1}{n} \right|$$

Obszar krytyczny:

$$k_0 = < d_n(1-\alpha), 1 >$$

gdzie wartość krytyczna  $d_n(1-\alpha)$  wyznaczana jest zgodnie z tablicą rozkładu Kołmogorowa-Lillieforsa jako  $\frac{0.886}{\sqrt{n}}$

Wyniki obliczeń z programu:

```
>d_stara_hala = 0.1297732
```

```
>wartosc_krytyczna_stara_hala = 0.1497613
```

Nie ma podstaw by odrzucić  $H_0$ . Wydajności pracy przy produkcji elementu na starej hali mają rozkład normalny.

```
>d_nowa_hala = 0.1060908
```

```
>wartosc_krytyczna_nowa_hala = 0.1278831
```

Nie ma podstaw by odrzucić  $H_0$ . Wydajności pracy przy produkcji elementu na nowej hali mają rozkład normalny.

3. Oszacować przedziałowo (współczynnik ufności 0,95) przeciętną wartość wydajności pracy przy produkcji elementu w starej hali. Obliczyć względną precyzję oszacowania i sprawdzić, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu starej hali.

a) Oszacowanie przedziałowo przeciętnej wartości wydajności pracy przy produkcji elementu w starej hali:

Wybór modelu:  $\sigma$  – nieznane,  $n > 30$

$$P\left(\bar{x} - u_\alpha \cdot \frac{s}{\sqrt{n}} < m < \bar{x} + u_\alpha \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Gdzie:

$m$  – szacowana średnia populacji

$u_\alpha$  – kwantyl rozkładu normalnego

$\alpha$  – wynosi 0,05

Otrzymany przedział:

$$P(33,52048 < m < 36,80638)$$

b) Obliczenie względnej precyzji oszacowania:

Względną precyzję oszacowania obliczamy ze wzoru:

$\delta = \frac{d}{\bar{x}} \cdot 100\%$ , gdzie  $d$  (bezwzględny błąd szacunku) jest wielkością odejmowaną i dodawaną do średniej  $\bar{x}$ .

$$d = u_\alpha \cdot \frac{s}{\sqrt{n}} = 1,642952$$

Zatem:

$$\delta = 4.672332\%$$

c) Sprawdzenie, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu starej hali:

Umowne przedziały miary precyzji względnej

$P = 1 - \alpha$	$\delta$	Wnioskowanie o szacowanym parametrze
<0,90;0,99>	<5%	Uprawnione i całkowicie bezpiecznie
	5-10%	Możliwe, ale z zalecaną ostrożnością
	>10%	Należy natychmiast przerwać

Tabela nr.1

Wnioskując z tabeli, mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu starej hali.

4. Oszacować przedziałowo (współczynnik ufności 0,95) odchylenie standardowa wydajności pracy przy produkcji elementu w nowej hali. Obliczyć względną precyzję oszacowania i sprawdzić, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu nowej hali.

a) Oszacowanie przedziałowo odchylenia standardowego wydajności pracy przy produkcji elementu w nowej hali:

Wybór modelu:  $\sigma$  – nieznane,  $n > 30$

$$P\left(\frac{s}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{u_\alpha}{\sqrt{2n}}}\right) = 1 - \alpha$$



Gdzie:

$\sigma$  – szacowane odchylenie standardowe populacji

$u_\alpha$  – kwantyl rozkładu normalnego

$\alpha$  – wynosi 0,05

Otrzymany przedział:

$$P(6,066787 < \sigma < 9,10097)$$

b) Obliczenie względnej precyzji oszacowania:

Względna precyzje oszacowania obliczamy ze wzoru:

$\delta = \frac{d}{s} \cdot 100\%$ , gdzie d (bezwzględny błąd szacunku) wynosi:

$$d = \frac{1}{2}(\sigma_2 - \sigma_1) = 1.517092$$

Gdzie:  $\sigma_1$ - początek przedziału,  $\sigma_2$ - koniec przedziału

Zatem:

$$\delta = 20,83803\%$$

c) Sprawdzenie, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu nowej hali:

Wnioskując z tabeli nr.1, nie mamy podstaw do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu nowej hali.

5. Czy na poziomie istotności 0,05 można stwierdzić, że wartości wydajności pracy przy produkcji elementu w starej hali są większe(sformułować i zweryfikować odpowiednią hipotezę)?

Aby dobrać odpowiednią statystykę do porównania średnich różnych populacji przeprowadziliśmy test Fishera-Snedecora.

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

gdzie  $\sigma_1^2$ ,  $\sigma_2^2$  to wariancje wydajności a nowej i starej hali.

Wyznaczenie statystyki testowej:

$$F = \frac{s_1^2}{s_2^2}$$

gdzie  $s_1$  – wariancja z próby dla nowej hali,  $s_2$  – wariancja z próby dla starej hali

Obszar krytyczny:

$$(F_{kr} ; \infty)$$

gdzie  $F_{kr}$  – wartość krytyczna dla poziomu istotności  $\alpha=0,05$  o stopniach swobody:  $u = 47$  i  $v=34$

```
>wartosc_testu_fishera = 2.155284  
>wartosc_krytyczna_rozkladu_fishera = 1.721589
```

Odrzucamy  $H_0$  na rzecz  $H_1$ . Przyjmujemy, że wariancje wydajności pracy na obu halach różnią się w sposób statystycznie znaczący, czyli dla dalszych obliczeń przyjmujemy statystykę Cochran-Coxa.

$$H_0: m_1 = m_2$$

$$H_1: m_2 > m_1$$

gdzie  $m_2$  – średnia wydajności pracy na starej hali  $m_1$  – średnia wydajności pracy na nowej hali

Wyznaczenie parametrów:

$$z_1 = \frac{s_1^2}{n_1-1} \qquad z_2 = \frac{s_2^2}{n_2-1}$$

Wyznaczenie statystyki:

$$C = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{z_1 + z_2}}$$

Obszar krytyczny:

$$(C_{kr}(1 - \alpha) ; \infty)$$

$$\text{gdzie } C_{kr}(1 - \alpha) = \frac{z_1 t_1(1-\alpha; n_1-1) + z_2 t_2(1-\alpha; n_2-1)}{z_1 + z_2}$$

Wartości  $t_1$  oraz  $t_2$  to wartości kwantyli rozkładu T-Studenta.

```
> c = 1.234277  
>wartosc_krytyczna_c = 1.683006
```

Nie ma podstaw by odrzucić  $H_0$ , czyli nie możemy stwierdzić, że wartości wydajności pracy przy produkcji elementu w starej hali są większe.