30.05.2018 Gliwice



**Metody Statystyczne**

Sprawozdanie z projektu 18

AEI, Informatyka sem. 4

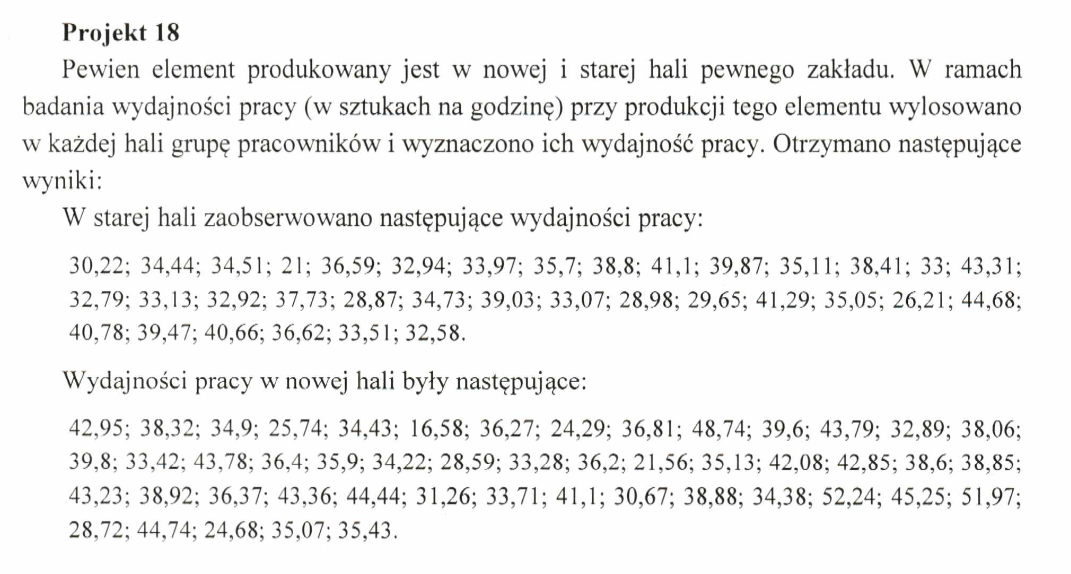
grupa 6, sekcja 5

Martyna Drabińska

Bartłomiej Krasoń

Michał Miciak

## Treść polecenia



## 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 postawie 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 | **mean** |
| Mediana | **median** |
| Kwartyl Q1 | **quantile** |
| Kwartyl Q3 | **quantile** |
| MIARY ZRÓŻNICOWANIA | |
| Wariancja nieobciążona | **var** |
| Wariancja obciążona | **var** |
| Odchylenie standardowe nieobciążone | **sd** |
| Odchylenie standardowe obciążone | **sd** |
| Odchylenie przeciętne od średniej | szcz_odch_przec_od_sr.PNG |
| Odchylenie przeciętne od mediany | szcz_odch_przec_od_me.PNG |
| Odchylenie ćwiartkowe | odch_cw.PNG |
| Współczynnik zmienności | wspol_zmienn.png |
| MIARY ASYMETRII | |
| Współczynnik skośności | s_M3.PNGA.PNG  , gdzie |
| MIARY KONCENTRACJI | |
| Kurtoza | s_M4.pngK.PNG  , gdzie |
| Exces | exces.PNG |

*\* - nazwy funkcji pogróbione*

Do wyznaczenia miar szeregów rozdzielczych skorzystaliśmy z następujących wzorów:

Ogólne oznaczenia:

- środek *i-tego* przedziału

- liczebność *i-tego* przedziału

|  |  |
| --- | --- |
| **MIARA** | **SPOSÓB WYZNACZENIA** |
| MIARY PRZECIĘTNE | |
| Średnia arytmetyczna | **R_srednia.PNG** |
| Moda | **r_moda.PNG**  – lewy koniec przedziału z modą  - długość przedziału z modą  - liczebnośc przedziału z modą  - liczebnośc przedziału poprzedzającego przedział z modą  - liczebnośc przedziału następującego po przedziale z modą |
| Mediana | **r_mediana.PNG**  - lewy koniec przedziału z medianą  - długość przedziału z medianą  - liczebność przedziału z medianą  - numer przedziału zawierającego medianę |

|  |  |
| --- | --- |
| Kwartyl Q1 | r_q1_q3.PNG  - lewy koniec przedziału zawierającego  - długość przedziału zawierającego  - liczebność przedziału zawierającego  - numer przedziału zawierającego |
| Kwartyl Q3 | **r_q1_q3.PNG**  - lewy koniec przedziału zawierającego  - długość przedziału zawierającego  - liczebność przedziału zawierającego  - numer przedziału zawierającego |
| MIARY ZRÓŻNICOWANIA | |
| Wariancja nieobciążona | **r_wariancja_nieobciazona.PNG** |
| Wariancja obciążona | r_wariancja.PNG |
| Odchylenie standardowe nieobciążone | **r_odch_stand_nieobciazone.PNG** |
| Odchylenie standardowe obciążone | r_odch_stand.PNG |
| Odchylenie przeciętne od średniej | r_odch_przec_od_sr.PNG |
| Odchylenie przeciętne od mediany | r_odch_przec_od_me.PNG |
| Odchylenie ćwiartkowe | odch_cw.PNG |
| Współczynnik zmienności | wspol_zmienn.png |
| MIARY ASYMETRII | |
| Wskaźnik asymetrii | wskaznik_asym.png |
| Współczynnik skośności | s_M3.PNGA.PNG  , gdzie |
| MIARY KONCENTRACJI | |
| Kurtoza | s_M4.pngK.PNG  , gdzie |
| Exces | exces.PNG |

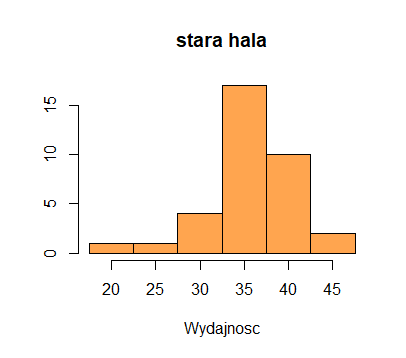
Obliczone miary danych starej hali:

|  |  |  |
| --- | --- | --- |
|  | szczegółowy | rozdzielczy |
| Średnia arytmetyczna | 35.16343 | 35.71429 |
| Moda |  | 35.75 |
| Mediana | 34.73 | 37.79412 |
| Kwartyl Q1 | 32.93 | 34.26471 |
| Kwartyl 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 |

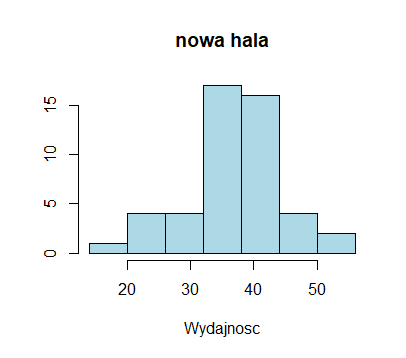
Obliczone miary danych nowej hali:

|  |  |  |
| --- | --- | --- |
|  | szczegółowy | rozdzielczy |
| Średnia arytmetyczna | 36.84271 | 36.875 |
| Moda |  | 37.57143 |
| Mediana | 36.385 | 37.29412 |
| Kwartyl Q1 | 33.6375 | 33.05882 |
| Kwartyl 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).

H0: Wydajności pracy przy produkcji elementu mają rozkład normalny.

H1: Wydajności pracy przy produkcji elementu nie mają rozkładu normalnego.

Wyznaczenie statystyki testowej:

gdzie:

Obszar krytyczny:

gdzie wartość krytyczna wyznaczana jest zgodnie z tablicą rozkładu Kołmogorowa-Lillieforsa jako

Wyniki obliczeń z programu:

>d\_stara\_hala = 0.1297732

>wartosc\_krytyczna\_stara\_hala = 0.1497613

Nie ma podstaw by odrzucić H0. 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 H0. 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 precyzje oszacowania i sprawdzić, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu starej hali.

1. Oszacowanie przedziałowo przeciętnej wartości wydajności pracy przy produkcji elementu w starej hali:

Wybór modelu: σ – nieznane , n>30

Gdzie:

m – szacowana średnia populacji

– kwantyl rozkładu normalnego

– wynosi 0,05

Otrzymany przedział:

1. Obliczenie względnej precyzji oszacowania:

Względną precyzję oszacowania obliczamy ze wzoru:

, gdzie d (bezwzględny błąd szacunku) jest wielkością odejmowaną i dodawaną do średniej .

Zatem:

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

Umowne przedziały miary precyzji względnej

|  |  |  |
| --- | --- | --- |
|  |  | 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 precyzje oszacowania i sprawdzić, czy mamy podstawy do uogólnienia otrzymanego przedziału ufności na całą populację wydajności pracy przy produkcji elementu nowej hali.

1. Oszacowanie przedziałowo odchylenia standardowego wydajności pracy przy produkcji elementu w nowej hali:

Wybór modelu: σ – nieznane , n>30

Gdzie:

– szacowane odchylenie standardowe populacji

– kwantyl rozkładu normalnego

– wynosi 0,05

Otrzymany przedział:

1. Obliczenie względnej precyzji oszacowania:

Względna precyzje oszacowania obliczamy ze wzoru:

, gdzie d (bezwzględny błąd szacunku) wynosi:

Gdzie: - początek przedziału, - koniec przedziału

Zatem:

1. 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.

H0:

H1:

gdzie to wariancje wydajności a nowej i starej hali.

Wyznaczenie statystyki testowej:

gdzie s1 – wariancja z próby dla nowej hali, s2 – wariancja z próby dla starej hali

Obszar krytyczny:

gdzie Fkr– wartość krytyczna dla poziomu istotności α=0,05 o stopniach swobody: u = 47 i v=34

>wartosc\_testu\_fishera = 2.155284

>wartosc\_krytyczna\_rozkladu\_fishera = 1.721589

Odrzucamy H0 na rzecz H1. Przyjmujemy, że wariancje wydajności pracy na obu halach różnią się w sposób statystycznie znaczący, czyli dla dalszych obliczeń przyjmujemy statystykę Cochrana-Coxa.

gdzie m2 – średnia wydajności pracy na starej hali m1 – średnia wydajności pracy na nowej hali

Wyznaczenie parametrów:

Wyznaczenie statystyki:

Obszar krytyczny:

gdzie

Wartości t1 oraz t2 to wartości kwantyli rozkładu T-Studenta.

> c = 1.234277

>wartosc\_krytyczna\_c = 1.683006

Nie ma podstaw by odrzucić H0, czyli nie możemy stwierdzić, że wartości wydajności pracy przy produkcji elementu w starej hali są większe.