Sprawozdanie z projektu 33 z metod statystycznych sem. IV

Natalia Mendera
Roch Fedorowicz
Daniel Cogiel
Sebastian Cyra
Jakub Włodarski

Czerwiec 2022
Politechnika Śląska

Spis treści

Zadanie 1: Miary szeregu szczegółowego

1.1 Treść zadania

Dokonać analizy zawartości cukru w procentach, w dostawach buraków cukrowych w województwie lubuskim i wielkopolskim, 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.

1.2 Opracowanie teoretyczne

1.2.1 Średnia próbkowa

1. Szereg szczegółowy

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.1}$$

n - liczebność próby
x_i - wartość elementu próby

2. Szereg przedziałowy

$$\bar{x} = \frac{\sum_{i=1}^{n} \dot{x}_i \cdot n_i}{\sum_{i=1}^{n} n_i} \tag{1.2}$$

n_i - liczebność przedziału
$\dot{x}_i$ - wartość środkowa przedziału

1.2.2 Mediana

1. Szereg szczegółowy

Mediana to wartość środkowa w uporządkowanym ciągu liczb. Jeśli liczebność próby jest nieparzysta, to medianą jest liczba o środkowym indeksie, w przeciwnym wypadku bierze się średnią arytmetyczną z dwóch środkowych liczb.

2. Szereg przedziałowy

$$m_e = x_0 + (N_{m_e} - n_{isk-1}) \cdot \frac{h_{m_e}}{n_{m_e}} \tag{1.3}$$

x_0 - dolna granica przedziału zawierającego medianę
$N_{me} = n \cdot 0,5$ - pozycja mediany
n_{isk-1} - liczebność skumulowana przedziału poprzedzającego liczebność skumulowaną mediany
h_{me} - szerokość przedziału zawierającego medianę
n_{me} - liczebność przedziału zawierającego medianę

1.2.3 Kwantyl

1. Szereg szczegółowy

   Kwantyl rzędu p stanowi taką wartość w próbie, że p wartości w próbie jest od niego mniejsza, a $1 - p$ większa.

2. Szereg przedziałowy

$$q = x_0 + (N_q - n_{isk-1}) \cdot \frac{h_q}{n_q} \tag{1.4}$$

x_0 - dolna granica przedziału zawierającego kwantyl
N_q - pozycja kwantyla (np. dla q_3: $N_q = n \cdot 0,75$)
n_{isk-1} - liczebność skumulowana przedziału poprzedzającego liczebność skumulowaną kwantyla
h_q - szerokość przedziału zawierającego kwantyla
n_q - liczebność przedziału zawierającego kwantyla

1.2.4 Moda (dominanta)

1. Szereg szczegółowy

   Dominanta to wartość najczęściej powtarzająca się w próbce.

2. Szereg przedziałowy

$$D_0 = x_0 + \frac{(n_0 - n_{-1})h_0}{(n_0 - n_{-1}) + (n_0 - n_{+1})} \tag{1.5}$$

x_0 - dolna granica przedziału dominującego
n_0 - liczebność przedziału dominującego
h_0 - długość przedziału dominującego
n_1 - liczebność przedziału poprzedzającego przedział dominujący
n_{+1} - liczebność przedziału następnego po przedziale dominującym

1.2.5 Zakres

$$r = x_{max} - x_{min} \tag{1.6}$$

1. Szereg szczegółowy
   x_{min} - element minimalny próby
   x_{max} - element maksymalny próby

2. Szereg przedziałowy
   x_{min} - dolna granica pierwszego przedziału szeregu
   x_{max} - górna granica ostatniego przedziału szeregu

1.2.6 Wariancja próbkowa

1. Estymator obciążony dla szeregu szczegółowego

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1.7}$$

2. Estymator nieobciążony dla szeregu szczegółowego

$$s_*^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1.8}$$

   Do równań 1.7 i 1.8:
   n - liczebność próby
   x_i - element próby
   $\bar{x}$ - średnia próbkowa szeregu

3. Estymator obciążony dla szeregu przedziałowego

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}n_i(\dot{x}_i - \bar{x})^2 \tag{1.9}$$

4. Estymator nieobciążony dla szeregu przedziałowego

$$s_*^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}n_i(\dot{x}_i - \bar{x})^2 \tag{1.10}$$

   Do równań 1.9 i 1.10:
   n - liczebność skumulowana szeregu
   n_i - liczebność przedziału
   $\dot{x}_i$ - wartość środkowa przedziału
   $\bar{x}$ - średnia próbkowa szeregu

1.2.7 Odchylenie standardowe

1. Estymator nieobciążony
$$\sigma_* = \sqrt{s_*^2} \tag{1.11}$$
2. Estymator obciążony
$$\sigma = \sqrt{s^2} \tag{1.12}$$

1.2.8 Odchylenie przeciętne

1. Od średniej dla szeregu szczegółowego
$$d_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \tag{1.13}$$
2. Od mediany dla szeregu szczegółowego
$$d_2 = \frac{1}{n}\sum_{i=1}^{n}|x_i - m_e| \tag{1.14}$$
   Do równań 1.13 i 1.14:
   n - liczebność próby
   x_i - element próby
   $\bar{x}$ - średnia próbkowa szeregu
   m_e - mediana szeregu
3. Od średniej dla szeregu przedziałowego
$$d_1 = \frac{1}{n}\sum_{i=1}^{n}n_i|\dot{x}_i - \bar{x}| \tag{1.15}$$
4. Od mediany dla szeregu przedziałowego
$$d_2 = \frac{1}{n}\sum_{i=1}^{n}n_i|\dot{x}_i - m_e| \tag{1.16}$$
   Do równań 1.15 i 1.16:
   n - liczebność skumulowana szeregu
   n_i - liczebność przedziału
   $\dot{x}_i$ - wartość środkowa przedziału
   $\bar{x}$ - średnia próbkowa szeregu
   m_e - mediana szeregu

1.2.9 Odchylenie ćwiartkowe

$$Q = \frac{q_3 - q_1}{2} \tag{1.17}$$

q_1 - kwartyl pierwszy dolny
q_3 - kwartyl trzeci górny

1.2.10 Klasyczny współczynnik zmienności

$$V_s = \frac{s_*}{\bar{x}} \tag{1.18}$$

s_* - estymator nieobciążony wariancji próbkowej szeregu
$\bar{x}$ - średnia próbkowa szeregu

1.2.11 Pozycyjny współczynnik zmienności

$$V_q = \frac{Q}{m_e} \tag{1.19}$$

Q - odchylenie ćwiartkowe szeregu
m_e - mediana szeregu

1.2.12 Momenty centralny (p-ty z kolei)

1. Szereg szczegółowy

$$M_p = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^p \tag{1.20}$$

   n - liczebność próby
   x_i - element próby
   $\bar{x}$ - średnia próbkowa szeregu

2. Szereg przedziałowy

$$M_p = \frac{1}{n}\sum_{i=1}^{n} n_i(\dot{x}_i - \bar{x})^p \tag{1.21}$$

   n - liczebność skumulowana szeregu
   n_i - liczebność przedziału
   $\dot{x}_i$ - wartość środkowa przedziału
   $\bar{x}$ - średnia próbkowa szeregu

1.2.13 Skośność

$$g_1 = \frac{M_3}{s^3} \tag{1.22}$$

M_3 - trzeci moment centralny szeregu
s - odchylenie standardowe próby

1.2.14 Kurtoza

$$krt = \frac{M_4}{s^4} \tag{1.23}$$

M_4 - czwarty moment centralny szeregu
s - odchylenie standardowe próby

1.2.15 Eksces

$$exc = krt - 3 \tag{1.24}$$

krt - kurtoza

1.3 Użyte funkcje

1.3.1 Szereg szczegółowy

Funkcje wbudowane

1. **mean()** - oblicza średnią próbkową
2. **median()** - oblicza medianę
3. **range()** - oblicza zakres
4. **var()** - oblicza wariancję próbkową z estymatorem nieobciążonym
5. **sd()** - oblicza odchylenie standardowe próby z estymatorem nieobciążonym
6. **mad()** - oblicza odchylenie przeciętne od mediany

Funkcje własne

1. **getmode()** - oblicza modę
2. **getQuantile()** - oblicza kwantyl podanego rzędu
3. **varO()** - oblicza wariancję próbkową z estymatorem obciążonym
4. **odchylenie_przecietne()** - oblicza odchylenie przeciętne od średniej
5. **skosnosc()** - oblicza skośność próby
6. **kurtoza()** - oblicza kurtozę próby

1.3.2 Szereg przedziałowy

Funkcje wbudowane

1. **round()** - zaokrągla wartość do danego miejsca po przecinku
2. **hist()** - tworzy histogram z zadanymi parametrami
3. **range()** - oblicza zakres

Funkcje własne

1. **sredniaRozdzielczy()** - oblicza średnią próbkową
2. **kwantylRozdzielczy()** - oblicza kwantyl podanego rzędu
3. **modaRozdzielczy()** - oblicza modę
4. **varNieobcRozdzielczy()** - oblicza wariancję próbkową z estymatorem nieobciążonym
5. **momentCentralnyRozdzielczy()** - oblicza moment centralny podanego rzędu
6. **wariancjaRozdzielczy()** - oblicza wariancję próbkową z estymatorem obciążonym
7. **odchylenieStdRozdzielczy()** - oblicza odchylenie standardowe próby z estymatorem obciążonym
8. **odchyleniePrzecietneRozdzielczy()** - oblicza odchylenie przeciętne od średniej
9. **madRozdzielczy()** - oblicza odchylenie przeciętne od mediany
10. **skosnoscRozdzielczy()** - oblicza skośność próby
11. **kurtozaRozdzielczy()** - oblicza kurtozę próby

1.4 Wyniki

1.4.1 Szereg szczegółowy

	Wartość	
Miara	**Lubuskie**	**Wielkopolskie**
Średnia	19.91706	18.55292
Mediana	20.02	18.605
Moda	24.15	19.3
Kwartyl 0.25	17.76	16.1
Kwartyl 0.75	22.7	20.72
Zakres	10.87 - 25.43	13.00 - 25.11
Wariancja	10.80879	7.993
Wariancja obc.	10.59685	7.826479
Odchylenie standardowe	3.287672	2.827189
Odchylenie standardowe obc.	3.255281	2.797584
Odchylenie przeciętne	2.635663	2.275833
Odchylenie przeciętne od mediany	3.899238	3.143112
Odchylenie ćwiartkowe	2.64	2.31
Klasyczny współczynnik zmienności	16.50682	15.23852
Pozycyjny współczynnik zmienności	0.8245163	0.8190549
Skośność	-0.3602522	-0.135047
Kurtoza	2.631021	2.399244
Eksces	-0.3689786	-0.6007556

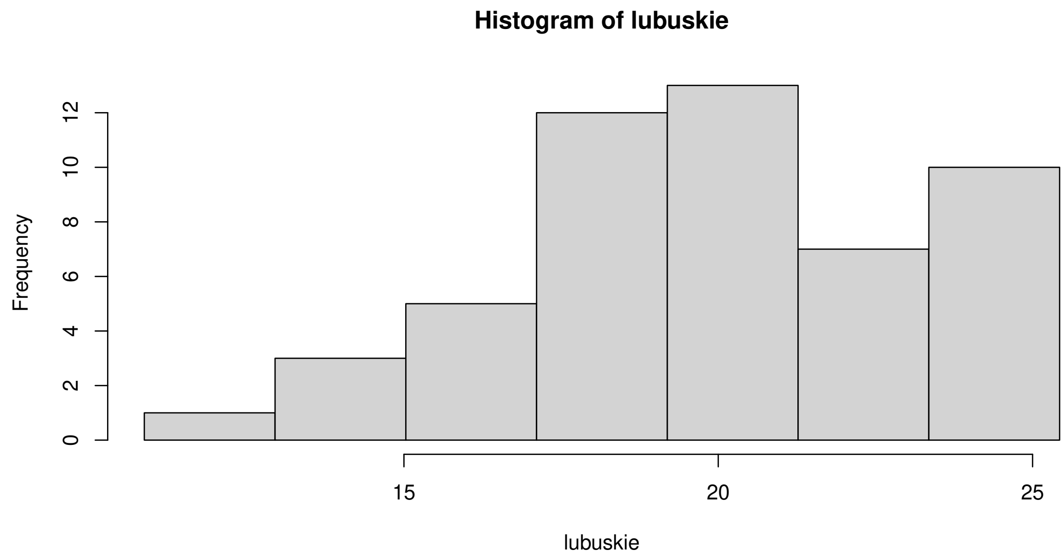
Tabela 1.1: Miary dla szeregu szczegółowego

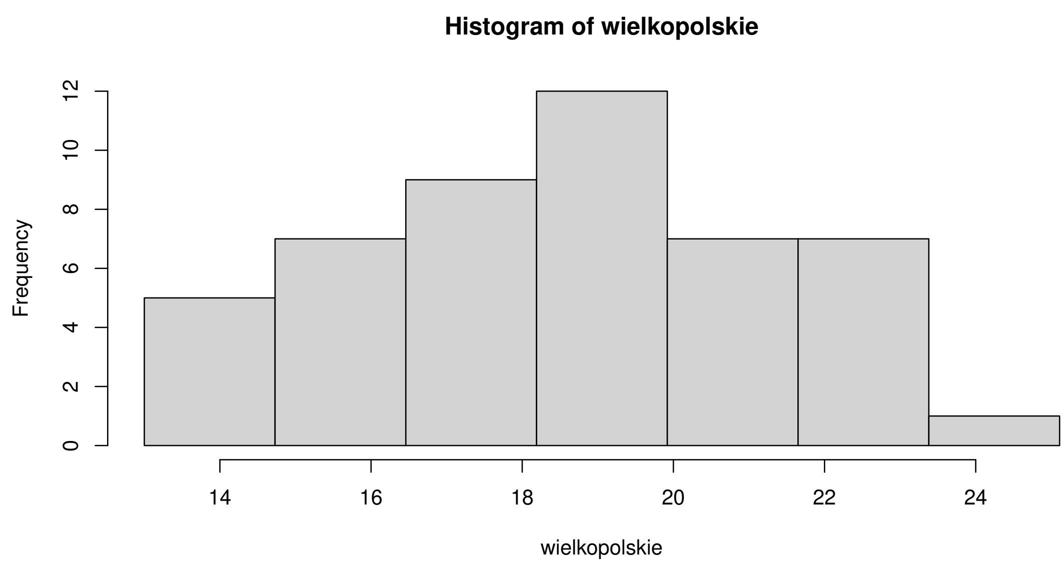
1.4.2 Szereg przedziałowy

	Wartość	
Miara	**Lubuskie**	**Wielkopolskie**
Średnia	19.90373	18.55042
Mediana	19.91	18.6225
Moda	19.48714	18.83875
Kwartyl 0.25	17.76	16.46
Kwartyl 0.75	22.53286	20.66143
Zakres	10.87 - 25.43	13.00 - 25.11
Wariancja	10.44783	7.763498
Wariancja obc.	10.24297	7.601758
Odchylenie standardowe	3.232311	2.786305
Odchylenie standardowe obc.	3.200464	2.757129
Odchylenie przeciętne	2.586205	2.297656
Odchylenie przeciętne od mediany	2.585098	2.288646
Odchylenie ćwiartkowe	2.386429	2.100714
Klasyczny współczynnik zmienności	16.23973	15.02018
Pozycyjny współczynnik zmienności	0.8156568	0.8065607
Skośność	-0.2851731	-0.01097241
Kurtoza	2.472798	2.134397
Eksces	-0.5272016	-0.8656031

Tabela 1.2: Miary dla szeregu przedziałowego

1.4.3 Histogramy





Zadanie 2: Test zgodności Kołmogorowa - Lillieforsa

2.1 Treść zadania

Sprawdzić, czy zawartości cukru w procentach, w dostawach buraków cukrowych mają rozkład normalny (test zgodności Kołmogorowa-Lillieforsa, współczynnik ufności 0,95).

2.2 Opis

Test Kołmogorowa-Smirnova z modyfikacją Lillieforsa, znany też jako test Kołmogorowa-Lillieforsa pozwala na zbadanie tezy, czy badana próba z populacji posiada rozkład normalny (lub jakikolwiek inny rozkład o znanej dystrybuancie). W celu sprawdzenia hipotezy zerowej zakładającej, że dana próba ma rozkład normalny, obliczana jest wartość statystyki będącej maksymalną różnicą pomiędzy wartością dystrybuanty rozkładu odniesienia ($F_0(x)$), a dystrybuantą empiryczną naszej próby ($\frac{i}{n}$). W danym punkcie pod uwagę brane są właściwie dwie różnice: progresywna ($|\frac{i}{n} - F_0(x)|$) oraz wsteczna ($|F_0(x) - \frac{i-1}{n}|$). Po obliczeniu maksymalnej z różnic w danym punkcie, obliczana jest maksymalna różnica we wszystkich punktach, dzięki czemu otrzymujemy wartość statystyki. Aby odrzucić lub stwierdzić, że nie mamy powodów do odrzucenia hipotezy zerowej wystarczy jeszcze policzyć przedział krytyczny, który możemy odczytać z tablic Kołmogorowa. Modyfikacja Lillieforsa natomiast upraszcza dobranie przedziału krytycznego dla testu poprzez modyfikacje tablic wartości krytycznej. Przykładowo, odczytując zmodyfikowane tablice możemy zauważyć, że przy przedziale $\alpha = 0.05$ i liczebności próby $n > 30$ możemy założyć, że dolny kres przedziału krytycznego to $\frac{0.886}{\sqrt{n}}$, co zostało wykorzystane w przeprowadzonym teście.

Hipoteza zerowa H_0 - zawartości cukru w procentach, w dostawach buraków cukrowych w danym województwie mają rozkład normalny

Hipoteza alternatywna H_1 - zawartości cukru w procentach, w dostawach buraków cukrowych w danym województwie nie mają rozkładu normalnego

2.3 Tabela

i	x_i	$\frac{i}{n}$	$F_0(x_i)$	$\lvert\frac{i}{n} - F_0(x_i)\rvert$	$\lvert\frac{i-1}{n} - F_0(x_i)\rvert$
1	x_1	$\frac{1}{n}$	$F_0(x_1)$	$\lvert\frac{1}{n} - F_0(x_1)\rvert$	$\lvert\frac{0}{n} - F_0(x_1)\rvert$
2	x_2	$\frac{2}{n}$	$F_0(x_2)$	$\lvert\frac{2}{n} - F_0(x_2)\rvert$	$\lvert\frac{1}{n} - F_0(x_2)\rvert$
...	...	...	...	...	...
				$max(\lvert\frac{i}{n} - F_0(x_i)\rvert)$	$max(\lvert\frac{i-1}{n} - F_0(x_i)\rvert)$

Tabela 2.1: Tabela Kołmogorowa - Lillieforsa

i - numer elementu szeregu pozycyjnego,
n - ilość elementów w całej populacji próby,
x_i - i-ty element szeregu pozycyjnego,
$\frac{i}{n}$ - dystrybuanta empiryczna dla x_i.

2.4 Wartość statystyki

$$D = max\left(max\left(\left|\frac{i}{n} - F_0(x_i)\right|\right),\ max\left(\left|\frac{i-1}{n} - F_0(x_i)\right|\right)\right) \tag{2.1}$$

2.5 Wyniki

2.5.1 Próba z województwa lubuskiego

1. Wartość statystyki
$$D = 0.08948465 \tag{2.2}$$

2. Warunek krytyczny odczytany z tablicy
$$\frac{0.886}{\sqrt{51}} = 0.1240648 \tag{2.3}$$

3. Wniosek

   Dla próby z województwa lubuskiego nie ma podstaw do odrzucenia hipotezy zerowej.

2.5.2 Próba z województwa wielkopolskiego

1. Wartość statystyki

$$D = 0.06001361 \tag{2.4}$$

2. Warunek krytyczny odczytany z tablicy

$$\frac{0.886}{\sqrt{48}} = 0.1278831 \tag{2.5}$$

3. Wniosek

   Dla próby z województwa wielkopolskiego nie ma podstaw do odrzucenia hipotezy zerowej.

Zadanie 3: Przedziałowa estymacja wartości średniej

3.1 Treść zadania

Oszacować przedziałowo (współczynnik ufności 0,95) przeciętną zawartości cukru w pro-centach, w dostawach buraków cukrowych w województwie lubuskim. Obliczyć względną precyzję oszacowania i sprawdzić, czy mamy podstawy do uogólniania otrzymanego przedziału ufności na całą populację zawartości cukru w dostawach buraków cukrowych, w województwie lubuskim.

3.2 Opis

Aby przedziałowo ocenić wartość średnią w próbie, należy skorzystać z twierdzenia o średniej próbkowej (3.1). Po stworzeniu odpowiedniego przedziału ufności (3.2) możemy, podstawiając do niego wzór na średnią próbkową, otrzymać przedział średniej próbkowej (3.3), który po odpowiednim przekształceniu, ze względu na wartość średnią (3.4) pozwoli na oszacowanie jej wartości przedziałowo w danej próbie.

3.3 Wzory

1. Twierdzenie o średniej próbkowej studentyzowanej

$$T = \frac{\bar{X} - m}{s_*}\sqrt{n} \tag{3.1}$$

$\bar{X}$ - średnia próbkowa otrzymana estymatorem,
m - faktyczna wartość średniej w populacji,
s_* - estymowana wartość odchylenia standardowego,
n - ilość elementów w próbie.

2. Przedział ufności ogólnie

$$P(t_1 < T < t_2),\ gdzie\ t_1 = t(\alpha_1),\ t_2 = t(1-\alpha_2) \tag{3.2}$$

Cecha ma rozkład normalny $N(m, \sigma)$ i nieznane jest odchylenie standardowe populacji, więc wykorzystano rozkład t-Studenta.
α_1, α_2 - przyjęte poziomy istotności,
$t_1(\alpha_1), t_2(1-\alpha_2)$ - kwantyle rozkładu t-Studenta o $n-1$ stopniach swobody.

3. Przedział ufności dla wartości średniej

$$P\left(t(\alpha_1) < \frac{\bar{X}-m}{s_*}\sqrt{n} < t(1-\alpha_2)\right) \tag{3.3}$$

4. Rozwiązując ze względu na m

$$\left(\bar{X} - t(1-\frac{\alpha}{2})\frac{s_*}{\sqrt{n}} < m < \bar{X} + t(1-\frac{\alpha}{2})\frac{s_*}{\sqrt{n}}\right) \tag{3.4}$$

uwzględniając, że:

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} \tag{3.5}$$

$$-t(\frac{\alpha}{2}) = t(1-\frac{\alpha}{2}) \tag{3.6}$$

5. Względna precyzja szacunku

$$prec = \frac{l_n}{2\bar{X}} \cdot 100\% \tag{3.7}$$

l_n - długość przedziału ufności
$\bar{X}$ - średnia próbkowa

3.4 Wyniki

1. Estymowane granice wartości średniej

$$(18.99239 < m < 20.84173) \tag{3.8}$$

2. Względna precyzja szacunku

$$prec = 4.642617\% \tag{3.9}$$

3. Wniosek
Uogólnienie wartości na cała populację jest całkowicie bezpieczne.

Zadanie 4: Przedziałowa estymacja odchylenia standardowego

4.1 Treść zadania

Oszacować przedziałowo (współczynnik ufności 0,95) odchylenie standardowe zawartości cukru w procentach, w dostawach buraków cukrowych w województwie wielkopolskim. Obliczyć względną precyzję oszacowania i sprawdzić, czy mamy podstawy do uogólniania otrzymanego przedziału ufności na całą populację zawartości cukru w dostawach buraków cukrowych, w województwie wielkopolskim.

4.2 Opis

Aby przedziałowo ocenić wartość odchylenia standardowego w próbie, należy skorzystać z twierdzenia o rozkładzie wariancji próbkowej (4.1). Po stworzeniu odpowiedniego przedziału ufności (4.2) możemy, podstawiając do niego wzór na rozkład wariancji próbkowej, otrzymać przedział rozkładu wariacji próbkowej (4.3), który po odpowiednim przekształceniu, ze względu na odchylenie standardowe (4.4) pozwoli na oszacowanie jego wartości przedziałowo w danej próbie.

4.3 Wzory

1. Twierdzenie o rozkładzie wariancji próbkowej

$$\chi^2 = \frac{(n-1)s_*^2}{\sigma^2} \tag{4.1}$$

χ^2 - statystyka chi-kwadrat,
s_*^2 - wariancja próbkowa,
n - ilość elementów w próbie.

2. Przedział ufności ogólnie

$$P(\chi_1^2 < \chi^2 < \chi_2^2),\ gdzie\ \chi_1^2 = \chi^2(\alpha_1, n-1),\ \chi_2^2 = \chi^2(1-\alpha_2, n-1) \tag{4.2}$$

$\chi^2(\alpha_1, n-1), \chi^2(1-\alpha_2, n-1)$ - kwantyle χ^2 o $n-1$ stopniach swobody rzędów odpowiednio: α_1 i $1-\alpha_2$.

3. Przedział ufności dla odchylenia standardowego

$$P\left(\chi^2(\alpha_1, n-1) < \frac{(n-1)s_*^2}{\sigma^2} < \chi^2(1-\alpha_2, n-1)\right) \tag{4.3}$$

4. Rozwiązując ze względu na σ

$$\left(\sqrt{\frac{(n-1)s_*^2}{\chi^2(1-\frac{\alpha}{2}, n-1)}} < \sigma < \sqrt{\frac{(n-1)s_*^2}{\chi^2(\frac{\alpha}{2}, n-1)}}\right) \tag{4.4}$$

uwzględniając, że:

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} \tag{4.5}$$

5. Względna precyzja szacunku

$$prec = \frac{l_n}{2s_*} \cdot 100\% \tag{4.6}$$

l_n - długość przedziału ufności
s_* - odchylenie standardowe próby

4.4 Wyniki

1. Estymowane granice odchylenia standardowego

$$(2.353547 < \sigma < 3.541282) \tag{4.7}$$

2. Względna precyzja szacunku

$$prec = 21.00558\% \tag{4.8}$$

3. Wniosek
   Uogólnienie wartości na cała populację jest całkowicie niepewne.

Zadanie 5: Test istotności różnicy średnich wartości

5.1 Treść zadania

Czy na poziomie istotności 0,05 można twierdzić, że zawartości cukru w procentach, w dostawach buraków cukrowych w województwie lubuskim są większe niż w województwie wielkopolskim (sformułować i zweryfikować odpowiednią hipotezę)?

5.2 Opis

Aby sprawdzić, czy wartość średniej populacji pierwszej jest większa, niż populacji drugiej, należy zacząć od odpowiedniego doboru testu oraz hipotez. Zaczynając od hipotezy, należy zastanowić się, jakie testy są możliwe do przeprowadzenia. W tym przypadku najlepiej sprawdzi się test istotności dla różnicy wartości średniej, gdzie **hipotezą zerową będzie równość wartości średniej w obu próbach**, natomiast **hipotezą alternatywną będzie, że wartość średnia wartości próby pierwszej jest większa niż wartość średnia próby drugiej**. Po ustaleniu hipotezy, pozostało wybrać odpowiednią statystykę testową. Z racji, iż należy zbadać wartości prób, których miary są w pełni estymowane z ich wartości (tj. nie znamy dokładnych ich wartości na przestrzeni populacji) do przeprowadzenia testu można wykorzystać dwie statystyki, a dobór ich zależy od równości wartości wariancji w porównywanych próbach. Z racji, że informację na temat równości wariancji w próbach należy otrzymać również poprzez test istotności, należy wpierw przeprowadzić test istotności dla różnicy wariancji, a dopiero potem w zależności od wniosków płynących z tego testu dobrać odpowiednią statystykę do głównego testu.

Aby więc sprawdzić równość wariancji w obu próbach, należy skorzystać ze statystyki F-Snedecora oraz twierdzenia o rozkładzie ilorazu dwóch wariancji próbkowych. Hipotezy w teście można postawić analogicznie do hipotez postawionych w teście głównym (tzn. **hipoteza zerowa zakłada równość, a hipoteza alternatywna, że wariancja w próbie pierwszej jest większa**). Obliczając wartość statystyki oraz porównując ją z przedziałem krytycznym dla danych zadanych w projekcie, można dojść do wniosku, że nie ma podstaw, aby odrzucić hipotezę alternatywną.

Co za tym idzie, opierając się na tym teście, można wnioskować, że wartości wariancji w próbach są równe i należy w doborze statystyki testowej dla testu głównego posłużyć się tym zakładającym równość wariancji w próbach. Oznacza to, że należy wykorzystać twierdzenie o rozkładzie dwóch średnich próbkowych dla równych wariancji. Odwołując się do postawionych na początku hipotezy zerowej i alternatywnej w teście głównym, aby udowodnić, że wartość średnia w próbie pierwszej jest większa niż w drugiej, należałoby odrzucić hipotezę zerową na korzyść alternatywnej. Aby móc to zrobić, należy oczywiście najpierw obliczyć wartość statystyki testowej oraz porównać jej wartość z przedziałem krytycznym. Po wykonaniu tych czynności dla danych zadanych w projekcie okazuje się, że właśnie tak też się dzieje, a co za tym idzie statystycznie można powiedzieć, że wartość średnia próby pierwszej jest większa od wartości średniej próby drugiej.

5.3 Wzory

1. Twierdzenie o rozkładzie ilorazu dwóch wariancji próbkowych

$$F = \frac{s_{*1}^2}{s_{*2}^2} \tag{5.1}$$

2. Hipoteza zerowa $H_0 : \sigma_1 = \sigma_2$
3. Hipoteza alternatywna $H_1 : \sigma_1 > \sigma_2$
4. Przedział krytyczny $K_0 = \langle F(1-\alpha, n_1-1, n_2-1), \infty)$
5. Twierdzenie o rozkładzie różnicy dwóch średnich próbkowych

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)\cdot s_{*1}^2 + (n_2-1)\cdot s_{*2}^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.2}$$

6. Hipoteza zerowa $H_0 : m_1 = m_2$
7. Hipoteza alternatywna $H_1 : m_1 > m_2$
8. Przedział krytyczny $K_0 = \langle t(1-\alpha, n_1+n_2-2), \infty)$

5.4 Wyniki

1. Wartość statystyki ilorazu dwóch wariancji próbkowych

$$F = 1.352282 \tag{5.3}$$

2. Przedział krytyczny K_0

$$K_0 = (1.614961, \infty) \tag{5.4}$$

3. Wniosek
   Nie ma podstaw, aby odrzucić hipotezę zerową, że odchylenia standardowe obu populacji są równe.

4. Wartość statystyki rozkładu różnicy dwóch średnich próbkowych

$$T = 2.207291 \tag{5.5}$$

5. Przedział krytyczny K_0

$$K_0 = (1.660715, \infty) \tag{5.6}$$

6. Wniosek
   Należy odrzucić hipotezę zerową na rzecz alternatywnej, że wartość średnia populacji pierwszej jest większa niż wartość średnia populacji drugiej.