Metody Statystyczne

kierunek: INF SSI

rok akademicki: 2021/2022

rok studiów: 2

semestr: 4

Projekt

prowadzący zajęcia: prof. dr hab. inż. Katarzyna Stąpor

sekcja 4-1

skład sekcji:

Wojciech Ptaś

Wojciech Siudy

Jan Kocurek

Tomasz Zawadzki

Bartosz Orlof

Paweł Skorupa

Gliwice, 31 maja 2022

Tematem projektu była analiza wyników finansowych firm handlowych zatrudniających do 5 pracowników funkcjonujących na terenie Wrocławia w 1997 r.

1. **Analiza struktury kosztów i obrotów**

Obliczono następujące miary położenia:

- Średnia $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Moda (najczęściej występująca wartość)
- Mediana $Q2$
- Kwartyl rzędu ¼ $Q1$
- Kwartyl rzędu ¾ $Q3$

Obliczono również miary zróżnicowania:

- Wariancja $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Odchylenie standardowe $s = \sqrt{s^2}$
- Średnie odchylenie bezwzględne $D = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- Współczynnik zmienności $\nu = \frac{s}{\bar{x}}$
- Rozstęp $x_{max} - x_{min}$
- Rozstęp ćwiartkowy $IQR = Q_3 - Q_1$

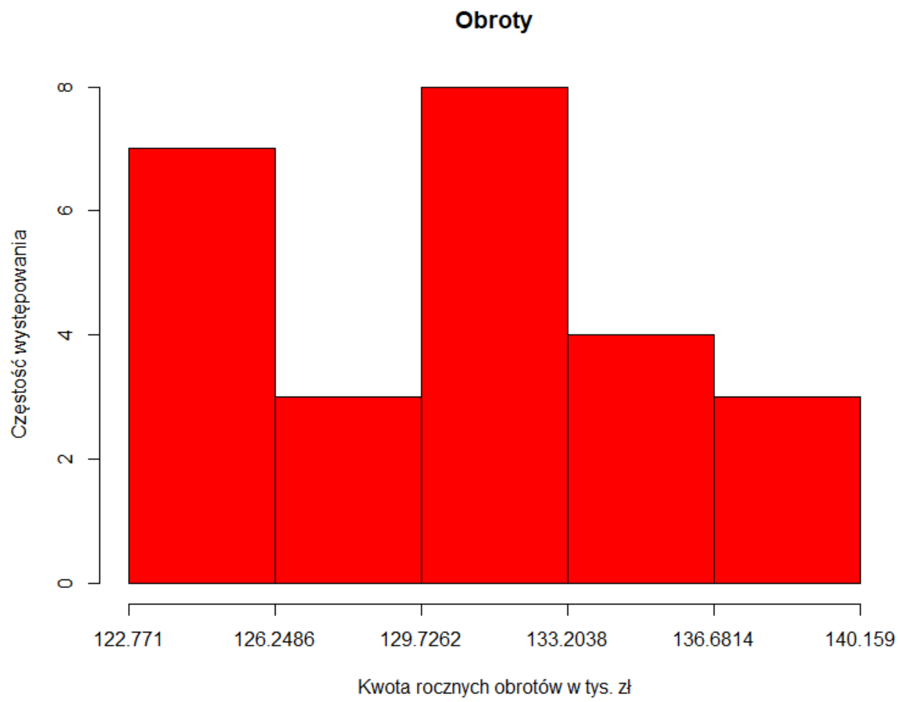
Następnie obliczono wybrane miary asymetrii:

- Współczynnik asymetrii (skośność) $A = \frac{\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|^3}{s^3}$

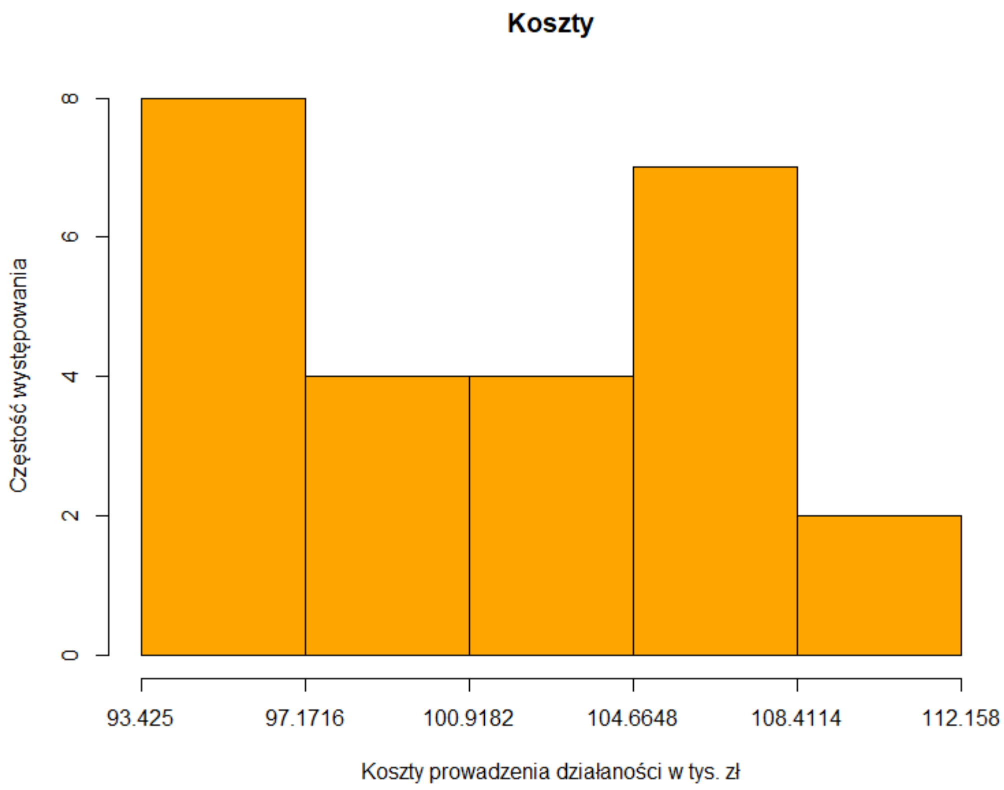
Obliczono również miary skupienia:

- Kurtoza $Kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|^4}{s^4}$

Rozkład obrotów wykazuje lekką prawostronną asymetrię (skośność jest równa 0,05) oraz spłaszczenie (kurtoza wynosi 2,1 < 3).



Rozkład kosztów również wykazuje prawostronną asymetrię (skośność jest w tym przypadku równa 0,18) oraz cechuje go spłaszczenie (kurtoza jest równa 1,93 < 3).



2. Sprawdzenie rozkładów kosztów i obrotów

W teście Kołmogorowa, na poziomie istotności testu α=0,05 zweryfikowano hipotezy, że rozkład kosztów i obrotów jest rozkładem normalnym. W tym celu skorzystano ze statystki testowej, będącej miarą rozbieżności między rozkładem hipotetycznym i empirycznym:

$$D_n = \max_x |F_0(x) - F_n(x)| = \max\,(d_n^-, d_n^+)$$

gdzie:

$$d_n^+ = \max\left|\frac{i}{n} - F_0(x)\right| \, oraz \; d_n^- = \max\left|F_0(x) - \frac{i-1}{n}\right| \; dla\; 1 \le i \le n$$

Zbudowano również obszar krytyczny, według następującego wzoru:

$$K_0 < d_n(1{-}\propto), 1 >,$$

gdzie $d_n(1{-}\propto)$jest wartością krytyczną odczytywaną z tablic dokładnego rozkładu Kołmogorowa. Dla obu populacji wartość ta była taka sama i wynosiła 0.2641

Wartość statystyki testowej dla obrotów wynosiła: 0,1273, a dla kosztów wynosiła: 0,1271.

Obliczone wartości nie należą do obszaru krytycznego, tak więc test Kołmogorowa nie daje podstaw do odrzucenia hipotez zerowych, mówiących o tym, że rozkład obrotów i kosztów jest rozkładem normalnym.

3. Szacunek przedziałowy kosztów

Próba jest niewielka, a odchylenie jest znana. Z tego powodu skorzystano z poniższego wzoru:

$$\left(\overline{x} - t\left(1-\frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n-1}}, \overline{x} + t\left(1-\frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n-1}}\right)$$

Obliczono również względną precyzję szacunku, korzystając ze wzoru:

$prec = \frac{d(\hat{a}_n)}{\hat{a}_n} * 100\%$, gdzie $d(\hat{a}_n) = \frac{l_n}{2}$

$l_n = a_g(X_1, \dots, X_n) - a_d(X_1, \dots, X_n)$

Otrzymano wartość $prec = 2{,}64\%$. Ponieważ wartość precyzji wynosi mniej niż 5%, oznacza to, że nasze wnioskowanie statystyczne jest całkowicie bezpieczne, więc mamy podstawy do uogólnienia wyników na całą populację małych firm.

4. Szacunek przedziałowy wariancji kosztów

Zważywszy na małą wielkość próby, skorzystano z rozkładu chi-kwadrat. Wykorzystany, w celu estymacji przedziałowej wariancji kosztów, wzór jest następujący:

$$\left(\frac{n * s^2}{\chi^2(1-\frac{\alpha}{2}, n-1)}, \frac{n * s^2}{\chi^2(\frac{\alpha}{2}, n-1)}\right)$$

Oszacowano przedziałowo wariancję kosztów, dla zadanego współczynnika ufności, a następnie obliczono względną precyzję szacunku, korzystając z tego samego wzoru, co w zadaniu 3. Ponieważ otrzymano wartość powyżej 10% (dokładnie $prec = 86{,}06\%$), wynik należy określić jako całkowicie niepewny, nie ma podstaw do uogólnienia go na całą populację.

5. Weryfikacja hipotezy o dochodowości branży

W celu dowiedzenia, czy branża jest dochodowa, przeprowadzono test istotności. Postawiono hipotezę zerową o równości wartości przeciętnych kosztów i obrotów:

$$H_0\colon m_o = m_k$$

którą spodziewano się odrzucić na rzecz następującej hipotezy alternatywnej:

$$H_1\colon m_o > m_k$$

Ponieważ nie posiadano informacji o tym, czy wariancje dwóch populacji są równe, przeprowadzono test, z tym samym współczynnikiem istotności.

Hipoteza zerowa: wariancje obu populacji są równe.

$$H_0\colon \sigma_o^2 = \sigma_k^2$$

Hipoteza alternatywna: wariancje obu populacji są różne

$$H_1\colon \sigma_o^2 \neq \sigma_k^2$$

Wartość statystyki testowej obliczono ze wzoru:

$$F = \frac{s_o^2}{s_k^2} = 0.85$$

A krawędź rozkładu odczytano z tablicy kwantyli rozkładu F Snedecora:

$$F(1-\alpha,\, n_o - 1, n_1 - 1) = 1.98$$

Test wykazał brak podstaw do odrzucenia hipotezy (na poziomie istności $\alpha = 0.05$) mówiącej, że wariancje dwóch populacji są równe.

Następnie zbudowano statystykę testową testu dotyczącego równości wartości oczekiwanych. Wartość statystyki testowej obliczono ze wzoru:

$$T = \frac{\overline{X_o} - \overline{X_k}}{\sqrt{\frac{n_o S_o^2 + n_k S_k^2}{n_o + n_k - 2} * \frac{n_o + n_k}{n_o * n_k}}} = 19.75$$

Wartość lewej krawędzi zbioru krytycznego odczytano z rozkładu kwantyli t-Studenta:

$$t(1-\alpha, n_o + n_k - 2) = 1.68$$

Wartość statystyki testowej należała do zbioru krytycznego, więc hipotezę zerową należało odrzucić na poziomie istotności $\alpha = 0.05$, na korzyść hipotezy alternatywnej, według której średnia wartość obrotów jest większa niż średnia wartość kosztów, a więc branża jest dochodowa.