

# 1. Podstawy pracy w PS IMAGO PRO

**Ćwiczenie 1.1.** Wczytaj dane z pliku *bankloan\_pl.sav* [6]. Jest to plik danych hipotetycznych zawierający informacje finansowe i demograficzne dotyczące 5000 klientów banku.

- a) Zapoznaj się ze zmiennymi dostępnymi w pliku oraz ich etykietami.
- b) Wypróbuj działanie ikony ETYKIETY WARTOŚCI.
- c) Przejdź do zakładki ZMIENNE i zwróć uwagę na ustawienie własności poszczególnych zmiennych. Które zmienne mają zdefiniowane wartości, a które braki danych?
- d) Jak zdefiniowano poziom pomiaru poszczególnych zmiennych?
- e) Wybierz z menu DANE → DEFINIUJ WŁAŚCIWOŚCI ZMIENNYCH... Przenieś wszystkie zmienne do pola ZMIENNE DO ZESKANOWANIA, odznacz sugerowane ograniczenia, a następnie przyjrzyj się kolejnym zmiennym. Sprawdź, czy mają one jakieś wartości, które można uznać za błędne lub nietypowe.

**Ćwiczenie 1.2.** Wczytaj dane z pliku *bankloan\_pl.sav*. Na podstawie dostępnych w nim zmiennych utwórz:

- a) zmienną *zadłużenie\_procent*, dodając do siebie zadłużenie na karcie kredytowej i inne zadłużenie, a następnie dzieląc sumę przez zmienną *dochód* i mnożąc przez 100,
- b) zmienną *równe* o wartościach 1, jeśli wartości zmiennych *zadłużenie\_stosunek* i *zadłużenie\_procent* są sobie równe, a 0, w przeciwnym przypadku,
- c) zmienną łańcuchową *decyzja* o szerokości 25, rekodując wartość 0 zmiennej *pożyczka* na napis *nie udzielono pożyczki*, a wartość 1 na napis *udzielono pożyczki*,
- d) zmienną numeryczną *decyzja\_num*, rekodując automatycznie zmienną *decyzja* (zwróć uwagę na to, jakie wartości ma nowa zmienna),
- e) zmienną *zadłużenie\_wysokość* o wartościach:
  - 1 (niskie), jeśli stosunek zadłużenia do dochodów ( $\times 100$ ) (zmienna *zadłużenie\_stosunek*) nie przekracza 5,
  - 2 (średnie), jeśli stosunek zadłużenia do dochodów mieści się w przedziale (5, 15] dolarów,

- 3 (wysokie), jeśli stosunek zadłużenia do dochodów mieści się w przedziale  $(15, 25]$  dolarów,
- 4 (bardzo wysokie), jeśli stosunek zadłużenia do dochodów przekracza 25.

f) zmienną *kategoria\_wiekowa* powstałą z kategoryzacji wieku na pięć przedziałów o zbliżonej liczebności.

## 2. Statystyka opisowa i grafika

**Ćwiczenie 2.1.** Otwórz plik *bankloan.pl.sav*. Wykonaj tabele częstości oraz wykres słupkowy prezentujący poziom wykształcenia klientów banku, a następnie tabelę częstości i wykres kołowy, na którym porównasz procent osób, którym udzielono i nie udzielono pożyczki.

**Ćwiczenie 2.2.** Ze zbioru danych *bankloan.pl.sav* wybierz tylko osoby młode, tj. w wieku poniżej 35 lat. Następnie porównaj rozkład zmiennej *pożyczka* (tabela częstości i wykres kołowy) w grupach wyróżnionych na podstawie wykształcenia.

**Ćwiczenie 2.3.** Otwórz plik *bankloan.pl.sav*. Wybierz do dalszej analizy osoby, których wiek nie przekroczył 40 lat. Wykonaj analizę statystyczną stosunku zadłużenia do dochodów w podziale na grupy osób, którym udzielono i nie udzielono pożyczki, następującymi trzema sposobami:

- ANALIZA → OPIS STATYSTYCZNY → CZĘSTOŚCI...,
- ANALIZA → OPIS STATYSTYCZNY → STATYSTYKI OPISOWE...,
- ANALIZA → OPIS STATYSTYCZNY → EKSPŁORACJA...

**Ćwiczenie 2.4.** Otwórz plik *bankloan.pl.sav*. Porównaj wartości zmiennych *zadłużenie\_karta* i *zadłużenie\_inne*, a następnie dokonaj standaryzacji zmiennych i porównaj je ponownie.

**Ćwiczenie 2.5.** Dla danych z pliku *bankloan.pl.sav* wykonaj następujące wykresy:

- zgrupowany wykres słupkowy prezentujący rozkład wykształcenia klientów banku z uwzględnieniem podziału na osoby, którym udzielono i nie udzielono pożyczki,

- b) zestawiony wykres słupkowy prezentujący rozkład wykształcenia klientów banku z uwzględnieniem podziału na osoby, którym udzielono i nie udzielono pożyczki,
- c) wykres słupkowy, na którym porównane zostaną średnie wartości dochodu gospodarstwa domowego w zależności od wykształcenia klienta,
- d) wykres skrzynkowy dochodów gospodarstwa domowego w podziale na grupy zależne od wykształcenia klienta,
- e) wykres liniowy prezentujący medianę dochodów gospodarstwa domowego w zależności od wieku klienta.

### 3. Przedziały ufności

**Ćwiczenie 3.1.** Zakładając, że dane z pliku *bankloan.pl.sav* pochodzą od osób, których można traktować jako prostą próbę losową wszystkich klientów pewnego banku (założenie to obowiązuje do końca skryptu), dla każdego poziomu wykształcenia z osobna wyznacz 95% i 99% przedziały ufności dla:

- a) stosunku zadłużenia do dochodów ( $\times 100$ ),
- b) odsetka osób, którym udzielono pożyczki.

### 4. Testowanie hipotez – jedna próba

**Ćwiczenie 4.1.** Generator liczb losowych wygenerował 20 liczb z rozkładu wykładniczego  $E(2)$ . Liczby są uporządkowane niemalejąco:

0,02	0,03	0,04	0,04	0,06	0,11	0,11
0,18	0,22	0,26	0,27	0,44	0,46	0,60
0,65	0,80	0,85	0,95	1,20	2,00	

Zilustruj badany rozkład na histogramie. Za pomocą testu Kołmogorowa na poziomie istotności 0,05 przetestuj zgodność tych danych z rozkładem  $E(2)$  (dane dostępne w pliku *wykladniczy.sav*).

**Ćwiczenie 4.2.** Otwórz plik *bankloan.pl.sav*. Wykonaj histogram zmiennej *Stosunek zadłużenia do dochodów ( $\times 100$ )* (*zadluzenie\_stosunek*). Zwróć uwagę na kształt rozkładu. Następnie wyznacz zmienną *pierwiastek* będącą pierwiastkiem kwadratowym stosunku zadłużenia do dochodów. Sprawdź, czy w każdej z grup wyznaczonych przez wykształcenie klientów banku zmienna *pierwiastek* ma rozkład normalny. Przyjmij poziom istotności 0,01.

**Ćwiczenie 4.3.** Dla danych z pliku *bankloan.pl.sav*, traktując je jako próbę prostą wszystkich klientów banku, zweryfikuj, przyjmując poziom istotności 0,05, poniższe hipotezy odnoszące się do całej populacji klientów tego banku.

- a) Co drugi klient, który ma ukończone studia podyplomowe, otrzymuje pożyczkę.
- b) W grupie klientów starszych (w wieku co najmniej 40 lat) pożyczkę otrzymuje co trzeci klient, który ma ukończone studia podyplomowe.

**Ćwiczenie 4.4.** Liczbę sprzedanych biletów MZK w Toruniu w kolejnych niedzielach maja i czerwca przedstawia tabelka.

Numer niedzieli	1	2	3	4	5	6	7	8
Liczba biletów w tys.	2,9	3,3	3,2	3,2	3,2	3,0	2,9	3,1

Na podstawie tych danych (dostępnych również w pliku *bilety.sav*), na poziomie istotności  $\alpha = 0,1$ , przetestuj hipotezę, że średnia liczba sprzedawanych biletów w niedzielę jest równa 3,2 tys. przeciw hipotezie, że średnia sprzedawanych biletów jest

- a) różna od 3,2 tys.,
- b) mniejsza niż 3,2 tys.,

jeżeli wiadomo, że liczba sprzedawanych biletów ma rozkład normalny.

**Ćwiczenie 4.5.** Dla danych z pliku *bankloan.pl.sav*, traktując je jako próbę prostą wszystkich klientów banku, używając testu  $t$  Studenta dla jednej próby przy poziomie istotności 0,05, zweryfikuj poniższe hipotezy odnoszące się do całej populacji klientów tego banku.

- a) W grupie osób, które ukończyły studia podyplomowe, przeciętne zadłużenie wynosi 10% wysokości dochodów.
- b) Co drugi klient, który ma ukończone studia podyplomowe, otrzymuje pożyczkę.

**Ćwiczenie 4.6.** Dla danych z pliku *bankloan.pl.sav*, traktując je jako próbę prostą wszystkich klientów banku, przy poziomie istotności 0,05, zweryfikuj hipotezę, że w grupie osób młodych (wiek nie więcej niż 35 lat), które ukończyły studia podyplomowe, przeciętny dochód gospodarstwa domowego wynosi 50 tys. dolarów.

## 5. Testowanie hipotez – dwie próby niezależne

**Ćwiczenie 5.1.** Otwórz plik *bankloan\_pl.sav*. Na podstawie zebranych w nim danych zweryfikuj hipotezy, przyjmując poziom istotności 0,05.

- a) Wybierz do analizy osoby dojrzałe (w wieku co najmniej 35 lat), które ukończyły studia podyplomowe. Zweryfikuj hipotezę, że grupy klientów, którym udzielono i nie udzielono pożyczki, różnią się ze względu na stosunek zadłużenia do dochodów.
- b) Wybierz do analizy osoby, które ukończyły studia wyższe. Zweryfikuj hipotezę, że przeciętny dochód gospodarstwa domowego osób młodych (w wieku poniżej 35 lat) jest niższy niż przeciętny dochód gospodarstwa domowego osób dojrzałych (w wieku co najmniej 35 lat).

**Ćwiczenie 5.2.** Otwórz plik *bankloan\_pl*. Przyjmując poziom istotności 0,05, zweryfikuj hipotezę, że grupy klientów, którym przyznano i nie przyznano pożyczki różnią się wysokością dochodów gospodarstwa domowego

- a) dla klientów, którzy ukończyli studia podyplomowe,
- b) dla klientów w wieku powyżej 50 lat, którzy nie ukończyli szkoły średniej.

## 6. Testowanie hipotez – dwie próby zależne

**Ćwiczenie 6.1.** Otwórz plik *bankloan\_pl.sav*. Zweryfikuj hipotezę o różnicy w wysokości zadłużenia na karcie kredytowej i innego rodzaju zadłużenia dla populacji osób:

- a) które ukończyły studia wyższe,
- b) w wieku powyżej 50 lat, które ukończyły studia wyższe.

**Ćwiczenie 6.2.** Otwórz plik *bankloan\_pl.sav*. Przyjmując poziom istotności 0,05, zweryfikuj hipotezę o różnicy w wysokości zadłużenia na karcie kredytowej i innego rodzaju zadłużenia dla populacji osób:

- a) w wieku poniżej 25 lat, które ukończyły studia wyższe,
- b) w wieku poniżej 24 lat, które ukończyły studia wyższe.

## 7. Badanie zależności zmiennych

**Ćwiczenie 7.1.** Na podstawie danych z pliku *bankloan.pl.sav* zweryfikuj hipotezę o zależności faktu udzielenia przez bank pożyczki i wykształcenia klienta.

**Ćwiczenie 7.2.** Otwórz plik *bankloan.pl.sav*. Utwórz zmienną o wartościach 1, gdy klient ma ponad 40 lat, i 0 w przeciwnym przypadku. Następnie wybierz klientów, którzy ukończyli studia podyplomowe. Sprawdź, czy decyzja o udzieleniu pożyczki i fakt, że klient ma ponad 40 lat, są niezależne.

**Ćwiczenie 7.3.** Otwórz plik *bankloan.pl.sav*. Wybierz do analizy klientów, którzy mają ponad 40 lat i ukończyli studia wyższe. Oceń jak silna jest w tej grupie zależność zadłużenia innego niż karta kredytowa i dochodów gospodarstwa domowego.

# Bibliografia

- [1] BIECEK P.: *Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych*. <http://www.biecek.pl/Eseje/> (dostęp z dnia 31.08.2025).
- [2] HARNETT D. L., SONI A. K.: *Statistical Methods for Business and Economics*. Addison-Wesley Publishing Company, 1991.
- [3] KORONACKI J., MIELNICZUK J.: *Statystyka dla studentów kierunków technicznych i przyrodniczych*. Warszawa, WNT, 2006.
- [4] REES D.G.: *Essential Statistics*. London, Chapman&Hall, 1995.
- [5] SHESKIN D.J.: *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, Chapman&Hall/CRC, 2011.
- [6] *bankloan-pl.sav* – plik przykładowy programu PS IMAGO PRO zamieszczony w katalogu *Predictive Solutions/PS IMAGO PRO/nr\_wersji/IBM/SPSS/Statistics/nr\_wersji/Samples* przetłumaczony na język polski (tłumaczenie własne).

