

STATYSTYKA
DLA STUDENTÓW III ROKU
MATEMATYKI EKONOMICZNEJ I MATEMATYKI Z INFORMATYKĄ
Laboratorium
Ćwiczenie 1

Dane:

- **census1p.sta**

5000 danych, uzyskanych z Biura Spisów Powszechnych USA¹. Zmienna **DOCHÓD** podaje dochód roczny (w tys. \$) pełnozatrudnionych białych robotnic w wieku 16-65 lat

- **census2p.sta**

400 danych, uzyskanych na podstawie zbioru **census1p.sta**. Kolumn (zmienna **DOCHK**) zawiera średnie z **k** losowo wybranych danych z **census1p.sta**. **k** = 1, 24, 10, 25, 50, 100, 625

- **census3p.sta**

400 danych, uzyskanych na podstawie zbioru **census1p.sta**. Kolumn (zmienna **DOCHKL**) zawiera średnie z **k** losowo wybranych logarytmów danych z **census1p.sta**. **k** = 1, 24, 10, 25, 50, 100, 625

Źródło danych: US Census, 1990

1. Import danych tekstowych Literatura [1] str 56

1. Przekopij dane **census?p.dat** oraz **census?p.sta** z kartoteki d:\prow\p-m3z1\labor do swojej kartoteki osobistej.
2. Uruchom program **STATISTICA** -> **PRZEŁĄCZNIK MODUŁÓW** -> **ZARZĄDZANIE DANYMI** -> **PLIK** -> **IMPORTUJ** -> **WIĘCEJ**
 - (a) wybierz ze swojej kartoteki plik **census1p.dat**,
 - (b) **OPCJE** -> **ZMIENNYCH 1** (BARDZO WAŻNE - JEST TU JEDNA ZMIENNA!)
 - (c) **SEPARATOR** - zestaw standardowy
 - (d) po wciśnięciu OK nie przejmuj się komunikatem, że jest za mało wierszy
 - (e) **ZAPISZ: TAK** zgodnie z sugerowanym przez komputer adresem
 - (f) Zrób to samo z plikiem **census2p.dat**, pamiętając, że jest tu 8 zmiennych

¹USA Census Bureau (census od łac. **censaere** - oceniać. Patrz cenzura, cenzor (w starożytnym Rzymie urzędnik przeprowadzający spisy powszechne))

2. Rozkład zmiennej losowej

1. PRZEŁĄCZNIK MODUŁÓW -> PODSTAWOWE STATYSTYKI I TABELĘ ([1] str 94)
2. Otwórz plik census1p.sta
3. Przejdź do opcji STATYSTYKI OPISOWE
4. Wybierz ZMIENNE->WSZYSTKIE
5. Zapoznaj się z różnymi formami przedstawienia graficznego danych:
HISTOGRAM, WYKRESY RAMKOWE (wszystkie), WYKRES PRAWDOPODOBIENSTWA NORMALNEGO²
6. Zwróć uwagę, że spośród wykresów ramkowych jedynie ciekawą informację uzyskasz z wykresu mediana/kwartyle/rozstęp
7. Co możesz ciekawego powiedzieć o rozkładzie zmiennej **DOCHÓD** histogramu i wykresu prawdopodobieństwa normalnego? Zwróć uwagę jak w obrazie graficznym ujawnia się prawostronna asymetria tych danych i co to w praktyce oznacza? Czy możesz powiedzieć, że zmienna **DOCHÓD** pochodzi z rozkładu normalnego?
8. Wykonaj obliczenia podstawowych parametrów statystycznych: średniej, odchylenia standardowego, mediany kwartyli:
 - (a) Powróć do panelu STATYSTYKI OPISOWE i wybierz wymienione parametry statystyczne z przycisku WIĘCEJ STATYSTYK a potem wykonaj obliczenia przyciskiem SZCZEGÓŁOWE STATYSTYKI OPISOWE
 - (b) Zapisz te wyniki w oknie wyników. Dane te będą dla nas populacją i do niej będziemy odnosić wszystkie późniejsze obliczenia.
 - (c) Zapisz te wyniki jako dane: PLIK -> ZAPISZ JAKO DANE: cen1st
 - (d) PRZEŁĄCZNIK MODUŁÓW -> ZARZĄDZANIE DANYMI OTWÓRZ DANE: cen1st
 - (e) dodaj dwie nowe zmienne **Youle** i **Bowley** w których obliczysz współczynniki asymetrii Youle'a i Bowley'a:

$$\frac{\text{średnia-mediana}}{\text{odchylenie standardowe}} \quad (\text{współczynnik Youle'a})$$
$$\frac{\text{kw górny} + \text{kw dolny} - 2 * \text{mediana}}{\text{kw górny} + \text{kw dolny}} \quad (\text{współczynnik Bowley'a})$$

Przedyskutuj wartości tych współczynników dla naszych danych. Skonfrontuj to z ich skrajnymi wartościami: -1,0,1. Co oznaczają te trzy wartości dla danych?

²POMOC: wykresy 2W -> wykresy statystyczne 2W

3. Centralne twierdzenie graniczne

1. Wykonaj wszystkie polecenia z punktu 2. Wyświetl na ekranie jednocześnie wszystkie histogramy, wykresy ramkowe i wykresy prawdopodobieństwa ustawiając opcję OKNA -> POZIOMO.
2. Zaobserwuj jak w miarę wzrostu liczby danych branych do średniej rośnie zgodność z rozkładem normalnym.
3. Po zapisaniu wyników z tabeli parametrów statystycznych (analogicznie punkt 3.8.c - 3.8.e) stwórz dwie nowe kolumny: jedną o nazwie **N** zawierającą liczby 1, 2 4, 10, 25, 50, 100, 625 i drugą o wartościach:

$$\text{odchylenie standardowe} \cdot \sqrt{N}$$

Co zauważasz? Jakie twierdzenie z rachunku prawdopodobieństwa ilustruje zauważony przez ciebie fakt?

3. Transformacja zmiennych

1. Powtórz wszystkie polecenia punktu 3. Co zauważasz?
2. Porównaj wykresy funkcji logarytm i wykres prawdopodobieństwa normalnego z punktu 2.5 Czy widzisz coś ciekawego? Czy to wyjaśnia potrzebę zastosowania przekształcenia logarytmicznego dla naszych danych?

Literatura

1. Stanisław, A., **Przystępny kurs statystyki**, Statsoft Polska, 1998

Andrzej Dąbrowski, 2.11.99