

STATYSTYKA OPISOWA (lab. 8)

ANALIZA KORELACJI POMIĘDZY CECHAMI LICZBOWYMI

Przykład 1 (Wybrane wskaźniki (Europa))

ANALIZA MACIERZE KORELACJI W PROGRAMIE STATISTICA

Przeanalizujemy zależność pomiędzy *PKB per capita* a *poziomem bezrobocia w roku 2020* w grupie państw europejskich. Ponieważ obie cechy mają charakter liczbowy, więc odpowiednim narzędziem do wykonania analizy będzie współczynnik korelacji.

Proszę teraz zapoznać się z podstawowymi informacjami o współczynniku korelacji zawartymi w przypisie¹.

- 1) Aby wyznaczyć wartość współczynnika korelacji z grupy *Statystyki podstawowe i tabele* wybieramy **Macierze korelacji**. Następnie dokonujemy wyboru zmiennych za pomocą przycisku **Dwie listy zmiennych**. Na pierwszej liście wskazujemy cechę, którą uznajemy bardziej za przyczynę niż skutek (tu proponujemy wybrać *PKB per capita (2020)*), zaś na drugiej liście cechę, której wartości mogą być uzależnione od pierwszej (wybieramy *Stopę bezrobocia (2020)*), zakładając że bogate państwa mają mniejsze problemy z sytuacją na rynku pracy). Oczywiście zmienne można wybrać też na odwrót – wyniki obliczeń będą identyczne, zaś jedyna różnica będzie dotyczyła układu tabeli i wykresu.
- 2) Wywołujemy wyniki za pomocą przycisku **Podsumowanie**. Proszę odczytać i podać wartość współczynnika korelacji: $r = \dots\dots\dots$, a następnie zinterpretować otrzymane wyniki (czy zależność istnieje, a jeżeli tak to jaka jest jej siła i kierunek):
.....
.....
- 3) Proszę wznowić analizę i zilustrować wyniki za pomocą wykresu rozrzutu (zakładka **Więcej, 2W Rozrzutu**). Można też wywołać wykres **Z nazwami przypadków**, na którym wszystkie kraje będą podpisane (przy dużej liczbie przypadków ten wykres jest raczej trudny do sformatowania, tak by jego wygląd był estetyczny i czytelny, ale można się z niego na przykład dowiedzieć, w których krajach badane cechy przyjmują wartości skrajne).
- 4) W analogiczny sposób proszę zbadać wpływ *Indeksu korupcji*² na trzy zmienne: *PKB per capita (2020)*, *Oczekiwany czas trwania życia mężczyzn (2020)* i *Stopę bezrobocia (2020)*.

W tabeli proszę zamieścić wartości współczynników korelacji i dokonać ich interpretacji.

Wybrane wskaźniki społeczno-gospodarcze (2020)	Indeks (braku) korupcji 2020	
	Współczynnik korelacji	Interpretacja wartości r
PKB per capita	$r = \dots\dots\dots$	
Oczekiwany czas trwania życia mężczyzn	$r = \dots\dots\dots$	
Stopa bezrobocia	$r = \dots\dots\dots$	

Który ze wskaźników rozwoju społeczno-gospodarczego był najmocniej związany z Indekssem (braku) korupcji, a który najslabiej? Czy kraje o wyższym poziomie korupcji mają niższe czy wyższe PKB per capita?

.....

¹ Współczynnik korelacji liniowej służy do badania siły zależności liniowej pomiędzy dwoma cechami liczbowymi i jest wskaźnikiem przyjmującym wartości z przedziału -1 do 1 . O sile korelacji świadczy wartość bezwzględna współczynnika a znak o jego kierunku. Tak więc, współczynniki korelacji $0,9$ czy $-0,9$ świadczą o tej samej (bardzo wysokiej) sile korelacji, choć wnioski wyciągane na ich podstawie będą przeciwstawne – w pierwszym przypadku wraz ze wzrostem wartości jednej cechy wartości drugiej też rosną, a w drugim przypadku spadają. Czasem przyjmuje się następującą skalę przymiotnikową, dotyczącą siły korelacji:

- $|r| < 0,3$ – brak korelacji;
- $0,3 \leq |r| < 0,5$ – słaba korelacja;
- $0,5 \leq |r| < 0,7$ – przeciętna korelacja;
- $0,7 \leq |r| < 0,9$ – silna korelacja;
- $0,9 \leq |r| < 1$ – bardzo silna korelacja;
- $|r| = 1$ – idealna korelacja.

² Indeks korupcji jest wskaźnikiem wyznaczanym corocznie dla większości państw świata. Interpretując wyniki należy pamiętać, że jest on skonstruowany w takim sposób, że wyższe wartości oznaczają mniejszą korupcję, więc jego poprawna nazwa powinna brzmieć Indeks braku korupcji. Indeks ten może przyjmować wartości od 0 do 100 pkt.

STATYSTYKA OPISOWA (lab. 8)

ANALIZA KORELACJI POMIĘDZY CECHAMI LICZBOWYMI

Przykład 2 (Efekty rehabilitacji)

WYZNACZANIE WSPÓLCZYNNIKA KORELACJI ZA POMOCĄ WYKRESU ROZRZUTU

Celem analizy będzie zbadanie zależności pomiędzy wiekiem pacjenta a wyjściowym poziomem sprawności oraz wiekiem a efektami rehabilitacji. Tym razem zamiast analizy korelacji wykorzystamy możliwość obliczenia współczynnika korelacji jako pomocniczej informacji na wykresie rozrzutu. Oto kolejne etapy rozwiązania zadania:

- W arkuszu danych dodajemy nową zmienną, nazywamy ją *Efekty rehabilitacji* i wyliczamy wartości za pomocą formuły (ma to być różnica między *sprawnością końcową* i *początkową*).
- Wybieramy polecenie *Wykresy / Wykresy 2W / Wykresy rozrzutu* i na pierwszej liście wskazujemy *Wiek* a na drugiej dwie zmienne: *Sprawność początkową* i *Efekty rehabilitacji*, następnie w zakładce *Więcej* zaznaczamy opcję *Korelacje* i wykonujemy wykresy.
- Proszę odczytać i zinterpretować korelację pomiędzy wiekiem i wyjściową sprawnością ($r = \dots\dots\dots$), a następnie wiekiem i efektami rehabilitacji ($r = \dots\dots\dots$).
- Jakie wnioski praktyczne można wysnuć z tych analiz, czy uprawniona jest teza, że „nie opłaca się” rehabilitować pacjentów w starszym wieku, bo uzyskują oni gorsze efekty rehabilitacji?

Przykład 3 (Informacje o krajach UE-28)

Celem analizy będzie zbadanie wpływu PKB (miernika bogactwa danego kraju) na oczekiwany czas trwania życia mężczyzn (miernik jakości życia). Pod uwagę weźmiemy dane z roku 2015. Sporządzając wykres rozrzutu na osi poziomej wskazujemy *PKB per capita 2015*, a na osi pionowej *Czas trwania życia mężczyzn 2015*. Robiąc wykres, proszę w zakładce *Więcej* zaznaczyć opcję *Korelacje*.

Z wykresu proszę odczytać i podać wartość współczynnika korelacji: $r = \dots\dots\dots$, a następnie zinterpretować otrzymane wyniki (czy zależność istnieje, a jeżeli tak to jaka jest jej siła i kierunek):

.....
.....

ANALIZA KORELACJI Z CZYNNIKIEM GRUPUJĄCYM

Na wykresie można zauważyć, dość wyraźnie wyodrębnione, dwie grupy państw o dłuższym i krótszym czasie trwania życia mężczyzn. Proszę wywołać wszystkie opcje wykresu i znaleźć zakładkę *Etykiety punktów*, gdzie należy zaznaczyć *Wyświetl etykiety punktów*. Nawet pobieżna analiza wyświetlonych etykiet każe przypuszczać, że poza aktualnym PKB na czas trwania życia mieszkańców kraju wpływa jego „historia” – grupa „gorszych” państw to wyłącznie państwa postkomunistyczne. Spróbujmy więc przeanalizować ponownie korelację pomiędzy PKB a czasem trwania życia z podziałem na dwie grupy państw.

Większość wykresów w programie *STATISTICA* można wykonać także w podziale na grupy, co znacząco uatrakcyjni i pogłębia te analizy. Potrzebujemy tylko zmiennej z opisem „typu” państwa. W tym celu skorzystamy z informacji zawartych w pliku *Wybrane wskaźniki (Europa)*, które przeniesiemy do pliku *Informacje o krajach UE-28*.

W tym celu proszę wykonać następujące czynności:

- Otworzymy plik *Informacje o krajach UE-28* oraz *Wybrane wskaźniki (Europa)*, a następnie wybieramy polecenie *Dane / Scal* gdzie jako *Plik 1* wskazujemy *Informacje o krajach UE-28*, a jako *Plik 2* – *Wybrane wskaźniki (Europa)*.
- W oknie *Opcje scalania* ustalimy tryb na *Według nazw przypadków*, zaznaczmy opcję *Usuń niezgodne przypadki* i przyciskiem *OK* wywołujemy nowy plik, który powstał z połączenia poprzednich – z tego pliku proszę usunąć wszystkie zmienne poza *Czas trwania życia mężczyzn 2015*, *PKB per capita 2015* oraz *Historia*.

Rozpoczynamy analizę (polecenia: *Wykresy / Wykresy 2W / Wykres rozrzutu*), gdzie należy wybrać zmienne tak samo jak na początku przykładu 3, zaznaczyć opcję *Korelacje*, a dodatkowo w zakładce *Skategoryzowane* proszę włączyć opcję *Kategoria X*, wskazać jako tę kategorię zmienną *Historia* oraz zaznaczyć opcję *Nalożone*. Należy tak sformatować wykres, by był zbliżony do prezentacji na rysunku.

Jak widać, nawet jeśli PKB w krajach postkomunistycznych jest na poziomie krajów kapitalistycznych, czas trwania życia jest o kilka lat krótszy. Ponadto w grupie państw kapitalistycznych różnice w bogactwie nie determinują czasu trwania życia ($r = 0,28$ – brak korelacji), a w grupie państw postkomunistycznych jest to istotne ($r = 0,50$ – przeciętna korelacja).

