

STATYSTYKA OPISOWA (lab. 2)

ARKUSZ DANYCH W PROGRAMIE STATISTICA

Przykład 1: Arkusz danych ankietowych

Proszę otworzyć plik *Efekty rehabilitacji*. Wiersze arkusza danych zawierają informacje o badanych jednostkach i są nazywane **PRZYPADKAMI**. W arkuszu danych ankietowych przypadki są anonimowe i nie mają nazw, tylko numery. Kolumny arkusza dotyczą cech analizowanych przypadków i określane są **ZMIENNYMI**. Właściwości zmiennej można edytować **po dwukrotnym kliknięciu jej nazwy**. Podczas pracy z arkuszem danych często trzeba odwoływać się do zmiennych – można to uczynić zarówno za pomocą ich nazw jak i numerów kolumn w arkuszu.

Globalne operacje na zmiennych i przypadkach można wykonywać za pomocą przycisków Zmienne ▾ Przypadki ▾ na pasku nad arkuszem danych.

Proszę udzielić odpowiedzi na następujące pytania:

- Ile osób (przypadków) liczy arkusz danych *Efekty rehabilitacji*:
- Ile zmiennych jest w tym arkuszu:

Przykład 2: Formuły w programie STATISTICA

Kontynuujemy pracę na pliku *Efekty rehabilitacji*. Naszym celem jest stworzenie **raportu dla NFZ**, w którym musimy opisać efekty rehabilitacji, które najpierw trzeba wyliczyć, jako różnicę pomiędzy wartościami z kolumny 5 i 4.

Za pomocą przycisku *Zmienne* i polecenia *Dodaj* proszę stworzyć nową zmienną na końcu arkusza i nazwać ją *Wielkość efektu rehabilitacji*, po czym wyliczyć za pomocą odpowiedniej formuły programu *STATISTICA*. Potrzebne informacje na temat formuł programu *STATISTICA* podano poniżej:

- formuły tworzy się za pomocą odwołania do zmiennych w postaci: $v + \text{numer zmiennej}$ (np. $v1$, $v10$, itd.);
- formuły wpisujemy **nie w komórce arkusza danych** (jak w *Excelu*), ale w tym samym oknie, w którym nadajemy nazwę zmiennym, w jego dolnej części;
- formuły muszą zaczynać się od symbolu „=”;
- formuły działają na całych kolumnach.

Korzystając z powyższych informacji proszę wyliczyć wartości zmiennej *Wielkość efektu rehabilitacji*.

Proszę wybrać polecenie *Statystyka / Statystyki podstawowe i tabele / Tabele liczości*, a następnie wybrać do analizy zmienną *Wielkość efektu rehabilitacji* i za pomocą przycisku *Podsumowanie* wywołać tabelę z informacją o liczbie pacjentów z określonymi efektami rehabilitacji. Zakładając, że terapia jest uzasadniona, gdy poprawa sprawności wyniesie **co najmniej 3 pkt**, proszę określić **u jakiego procenta chorych uzyskane wyniki uzasadniały podjęcie leczenia?**

Przykład 3: Arkusz danych przekrojowych

Proszę otworzyć plik danych *Informacje o krajach UE-28*. Jest to przykład tzw. danych przekrojowych. Podstawowa różnica w stosunku do danych ankietowych polega na tym, że przypadki są rozróżnialne (nie anonimowe) i mają swoje nazwy. Dane przekrojowe mają często (często, ale nie zawsze!) charakter **pełny** (to też odróżnia je od danych ankietowych), to znaczy dotyczą **wszystkich** województw, powiatów, gmin, państw, firm, a nie tylko ich losowo wybranej próby.

W arkuszu danych proszę odnaleźć i uzupełnić następujące informacje:

- poziom bezrobocia w Polsce w roku **2000**: i **2021**:
- poziom bezrobocia w Hiszpanii w roku **2000**: i **2021**:

Co można powiedzieć o sytuacji na rynku pracy w tych dwóch krajach?

Uwaga: arkusz danych przekrojowych można (i niejednokrotnie trzeba) sortować, jeśli celem analizy jest prezentacja

rankingu przypadków według pewnej cechy. Polecenie sortowania jest dostępne na pasku w postaci przycisku Z↓.

Proszę posortować dane malejąco lub rosnąco (zgodnie z treścią polecenia) według odpowiednich zmiennych i udzielić odpowiedzi na następujące pytania:

- najwyższe bezrobocie w **2024** roku odnotowano w i wynosiło ono%
- pod względem poziomu bezrobocia w **2024** roku Polska była na miejscu wśród krajów UE z poziomem bezrobocia%
- najwyższy poziom wskaźnika dzietności w **2021** roku zanotowano w i wynosił on , a najniższy w i wynosił on , Polska znajdowała się w tym rankingu na miejscu z poziomem dzietności

STATYSTYKA OPISOWA (lab. 2)

ARKUSZ DANYCH W PROGRAMIE STATISTICA

Przykład 4: Przykład arkusza danych czasowych

Proszę otworzyć plik danych *Transport w Polsce 1990-2024 (R)*. Jest to plik danych czasowych. Jego specyfika to nie tylko rozróżnialne przypadki, którymi są jednostki czasowe, ale także fakt, iż muszą one być ułożone w odpowiedniej kolejności.

Dane czasowe najczęściej prezentujemy w postaci graficznej – proszę wybrać polecenie WYKRESY / WYKRESY 2W / WYKRESY LINIOWE (ZMIENNE) i sporządzić wykres liczby samochodów osobowych w Polsce w latach 1990-2024. Proszę spróbować opisać najważniejsze spostrzeżenia dotyczące rozwoju motoryzacji indywidualnej w Polsce po 1990 r.:

UWAGA! Danych czasowych nie powinno się sortować względem wartości zmiennych – są one ułożone chronologicznie!

Przykład 5: Tworzenie i wypełnianie arkusza danych ankietowych

Załącznik do poniższych materiałów stanowi **kwestionariusz bardzo prostego badania ankietowego**. Poniżej opisano najważniejsze etapy tworzenia arkusza, w którym będzie można zawrzeć dane z przeprowadzonego badania.

1. Tworzenie nowego arkusza – polecenie *Plik / Nowy*
2. Zapisanie arkusza w „swoim” folderze pod nazwą *Przykładowa ankieta* – co pewien czas proszę zapisywać postępy swojej pracy.
3. Podczas tworzenia nowego arkusza, należy podać wyjściową liczbę zmiennych i przypadków – liczbę zmiennych ustalamy na podstawie analizy kwestionariusza ankiety, zaś liczbę przypadków na podstawie zakresu badania – proszę kwestię ilości zmiennych i przypadków przedyskutować z osobą prowadzącą zajęcia.
4. Nadanie nazw poszczególnym zmiennym/kolumnom arkusza – nazwy te powinny odpowiadać treści pytań, powinny być zwięzłe, ale jednocześnie precyzyjne (a więc kolejne zmienne będą się nazywać: *Płeć, Rodzaj studiów, Wiek, Miejsce zamieszkania*, etc.)
5. Aby zakodować odpowiedzi z pyt. 1 z drugiej części ankiety należy wykorzystać 6 kolumn – nazwy powinny odpowiadać treści składowych pytań (np. *Czystość i estetyka*, etc.):
6. Po nazwaniu wszystkich zmiennych proszę zaznaczyć cały arkusz, a następnie za pomocą poleceń *Format / Zmienne / Autodopasowanie* dopasować szerokość kolumn, tak by wprowadzone nazwy były dobrze widoczne
7. W przypadku, kiedy zmienna ma charakter nominalny (tekstowy) należy zakodować **wszystkie możliwe warianty odpowiedzi**, wykorzystując polecenie **ETYKIETY TEKSTOWE** dostępne w oknie opisu zmiennej – dla przykładu, zakodujemy warianty zmiennej *Płeć* jako: 1 – *kobieta*, 2 – *mężczyzna*; warianty zmiennej *Rodzaj studiów* jako: 1 – *stacjonarne*, 2 – *niestacjonarne* i tak dalej.
8. Odpowiedzi na pytanie 1 z drugiej części ankiety będą kodowane w następujący sposób: 1 – *bardzo niska*, 2 – *niska*, 3 – *średnia*, 4 – *dobra*, 5 – *bardzo dobra*.
9. W przypadku powtarzania się identycznych etykiet tekstowych dla kilku kolumn można je skopiować i wkleić. Można je też powielić za pomocą polecenia *Zastosuj do...* dostępnego w oknie edytora etykiet tekstowych.
10. Teraz można już wprowadzać dane stosując następujące reguły: dane wprowadzamy wierszami (wpisujemy ankietę od początku do końca i dopiero bierzemy kolejną), podczas wprowadzania danych używamy **wyłącznie** klawiszy z cyframi (0, 1, ..., 9), do kolejnej komórki przechodzimy za pomocą klawisza *Tab* lub →

Tworzenie arkusza

Nazywanie zmiennych

Etykiety tekstowe

Lista umiejętności, które powinny być opanowane po dzisiejszych zajęciach:

1. Otwieranie wskazanego pliku danych.
2. Umiejętność wskazania przypadków i zmiennych arkusza danych, odczytania ich liczby.
3. Umiejętność zmiany rozmiaru arkusza danych: dodawanie i usuwanie zmiennych lub przypadków.
4. Wylizania wartości nowych zmiennych na podstawie wartości zmiennych już istniejących za pomocą formuł programu *STATISTICA*.
5. Tworzenia arkusza danych do prostego kwestionariusza ankietowego.
6. Nazywanie zmiennych
7. Kodowanie odpowiedzi w postaci etykiet tekstowych.
8. Wpisywanie danych ankietowych.
9. Zapisywanie pliku danych.