

## Edycja, kodowanie i tabulacja danych



# Informacje wstępne

Szanowni Państwo!

W tej chwili wiele badań przeprowadza się on-line i wtedy pomija się bardzo **praco-** i **czasochłonny** etap kodowania i wprowadzania danych z wypełnionych kwestionariuszy.

Natomiast należy pamiętać, że **badania on-line** możemy przeprowadzić **tylko wśród użytkowników Internetu**.

Pomimo, że jest ich coraz więcej, należy pamiętać, że **najstarsze pokolenia** stosunkowo rzadko korzystają z Internetu.

Czasami, aby badania były rzetelne, **trzeba zbadać tę właśnie grupę**.

Ktoś mógłby powiedzieć, to wysyłamy tam ankieterów, którzy od razu zaznaczą odpowiedzi respondentów.

A co, jeśli tematyka badania jest **drażliwa**, a jest to produkt czy usługa skierowana **stricte do tej grupy**?

Struktura procentowa populacji ze względu na wyróżnione cechy					
w procentach					
	Miasto		Wieś		Razem
	Kobiety	Mężczyźni	Kobiety	Mężczyźni	
<b>15-29</b>	5,15%	5,29%	4,28%	4,55%	<b>19,27%</b>
<b>30-44</b>	8,64%	8,62%	5,28%	5,64%	<b>28,18%</b>
<b>45-59</b>	6,82%	6,36%	4,51%	4,75%	<b>22,44%</b>
<b>60-74</b>	8,03%	6,11%	3,91%	3,66%	<b>21,71%</b>
<b>75 i więcej</b>	3,60%	1,86%	1,95%	1,00%	<b>8,41%</b>
<b>Razem</b>	<b>32,24%</b>	<b>28,24%</b>	<b>19,93%</b>	<b>19,59%</b>	<b>100,00%</b>

# Informacje wstępne

Dlatego wtedy (ale nie tylko wtedy) **przeprowadza się badania**, gdzie **kwestionariusze występują w tradycyjnej (papierowej) formie**.

Te kwestionariusze wymagają najpierw rzetelnej edycji (redakcji), a następnie wprowadzenia do komputera.

**Dane zazwyczaj wprowadza się do arkusza kalkulacyjnego** (np. Excell-a), dlatego, że robi się to w nich najszybciej. Za pomocą arkusza kalkulacyjnego można wiele rzeczy policzyć.

Jednak **bardziej zaawansowane analizy** wymagają **użycia programów statystycznych**, takich jak np. **STATISTICA, SPSS, ...**

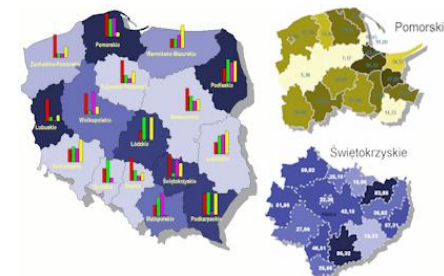
Z reguły importuje się do nich dane wprowadzone w Excell-u.

# Przetwarzanie danych

Aby otrzymane w wyniku badan empirycznych dane surowe mogły stać się podstawą podejmowania decyzji w przedsiębiorstwie, wymagają dalszego przetworzenia.

Przetworzenie to polega na:

- edycji danych,
- kodowaniu,
- tworzeniu nowych zmiennych (agregacji),
- tabulacji (opracowaniu zbiorczych tabel),
- wizualizacji danych w formie graficznej (wykresów, rysunków).



# Edycja danych

**Edycja danych** (zwana też **redagowaniem danych**) - jest procesem polegającym na dokładnym sprawdzeniu otrzymanych narzędzi badawczych (kwestionariuszy ankiet i kwestionariuszy wywiadów) pod względem:

- czytelności danych,
- kompletności,
- spójności i prawdziwości odpowiedzi
- oraz ich adekwatności.

**Cel edycji danych:** Upewnienie się, że dane są wiarygodne i gotowe do przetworzenia, wykluczając błędne lub nieadekwatne odpowiedzi.

## Edycja danych

W przypadkach, kiedy brakująca odpowiedź wynika z odpowiedzi na inne pytania, bądź możliwe jest określenie wielkości szacunkowej w danym pytaniu, badacz może **sam uzupełnić braki**, wpisując odpowiedź (najczęściej innym kolorem, niż wpisy oryginalne).

Jednak należy być bardzo ostrożnym we wprowadzaniu uzupełnień i poprawek, ponieważ nigdy nie można być pewnym intencji respondenta!

## Edycja danych

W sytuacji, kiedy nie ma możliwości dokonania interpretacji i uzupełnienia odpowiedzi na pytanie, wprowadza się dodatkową zmienną (**brak odpowiedzi**) lub eliminuje się kwestionariusz z dalszej analizy.

Kiedy w wielu kwestionariuszach brakuje odpowiedzi na konkretne pytanie lub są one udzielane w niewłaściwy sposób, to można wyeliminować z dalszej analizy **to** właśnie pytanie.

## Edycja danych

Edycja danych pozwala na uporządkowanie zebranego materiału i umożliwienie dokonania procesu kodowania danych.

Dokonanie edycji ma na celu taką selekcję danych, w efekcie której otrzyma się dane:

- spójne,
- trafne,
- istotne dla dalszej analizy,
- jednorodne
- oraz umożliwiający przeprowadzenie dalszego procesu redukcji i analizy (kodowania i tabulacji)

Ważne jest też, aby na tym etapie ponumerować kwestionariusze!!!

# Kodowanie danych

Aby móc dokonać nawet najprostszych operacji statystycznych „redukujących” różnorodność pierwotnych danych surowych, należy przełożyć odpowiedzi respondentów na ciąg symboli (najczęściej cyfr).

Kodowanie jest procedurą techniczną pomagającą w kategoryzacji danych.

**Cel kodowania:** Przekształcenie jakościowych danych w formę liczbową, która może być przetwarzana statystycznie.



# Procedura kodowania

Poprzez kodowanie surowe dane są przekształcane w symbole (kwantyfikatory) – zwykle liczbowe – które mogą być poddane tabelaryzacji (tabulacji) i zliczaniu.

Zakodowane dane wpisuje się do arkusza kalkulacyjnego.

- wiersze (**przypadki** – numery kolejnych kwestionariuszy)
- kolumny (**zmienne** – kolejne pytania lub ich części lub połączenia)

**Procedura kodowania jest różna, w zależności od charakteru pytań występujących w kwestionariuszu.**

# Procedura kodowania

zmienne



Dane: kod arytmetyczny.STA 6v \* 10c

LICZBOWE WARTOŚCI	1 PYT_17	2 PYT_18	3 PYT_19	4 PYT_20	5 PYT_21	6 PYT_22
ankieta 01	1	2	3	3	1	1
ankieta 02	0					2
ankieta 03	1	2	3	3	1	2
ankieta 04	0					1
ankieta 05	0					1
ankieta 06	0					4
ankieta 07	0					3
ankieta 08	0					2
ankieta 09	1	2	3	3	1	3
ankieta 10	0					1

przypadki





# Kodowanie pytań zamkniętych

Kodowanie odpowiedzi na pytania zamknięte może być przeprowadzone w każdej fazie realizacji badań.

W przypadku pytań zamkniętych kodowanie polega na przyporządkowaniu każdemu wariantowi odpowiedzi danego symbolu.

Jednakże w zależności od tego czy są to pytania z kafeterią dysjunktywną (możliwość wyboru tylko jednej odpowiedzi) czy też koniunktywną (możliwość wyboru więcej niż jednej odpowiedzi) stosuje się **różne sposoby kodowania**.

W przypadku tych pierwszych stosujemy **kod arytmetyczny** (natomiast dla zmiennych dychotomicznych możemy alternatywnie zastosować **kod binarny**), natomiast w przypadku tych drugich **kod geometryczny** lub/i **binarny (zero-jedynkowy)**.

# Kodowanie pytań zamkniętych z możliwością wyboru jednego wariantu odpowiedzi (**kod arytmetyczny**) (1)

**Pyt. 12. Czy jest Pani zadowolona z dezodorantu Nivea ?**

- zdecydowanie tak → 1
- raczej tak → 2
- raczej nie → 3
- zdecydowanie nie → 4
- jeszcze nie mam wyrobionej opinii → 5

**Numery  
poszczególnych  
kwestionariuszy  
to PRZYPADKI  
(w wierszach),  
natomiast  
kolejne pytania  
lub ich części lub  
połączenia to  
ZMIENNE  
(w kolumnach)!**

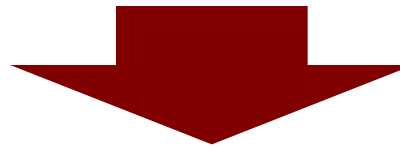
LICZBOWE WARTOŚCI	1 PYT_9	2 PYT_10	3 PYT_11	4 PYT_12	5 PYT_13	6 PYT_14
ank 01				1		
ank 02				2		
ank 03				5		
ank 04				2		
ank 05				3		
ank 06				2		
ank 07				1		
ank 08				4		
ank 09				4		
ank 10				2		



# Kodowanie pytań zamkniętych z możliwością wyboru jednego wariantu odpowiedzi dla zm. dychotomicznych (**kod binarny**)(2)

**17. Czy w ostatnim roku kupił(a) Pan(i) samochód ?**

- TAK → proszę przejść do pytania 18 → **1**
- NIE → proszę przejść do pytania 22 → 0



Dane: kod arytmetyczny.STA 6v \* 10c

LICZBOWE WARTOŚCI	1 PYT_17	2 PYT_18	3 PYT_19	4 PYT_20	5 PYT_21	6 PYT_22
ankieta 01	1	2	3	3	1	1
ankieta 02	0					2
ankieta 03	1	2	3	3	1	2
ankieta 04	0					1
ankieta 05	0					1
ankieta 06	0					4
ankieta 07	0					3
ankieta 08	0					2
ankieta 09	1	2	3	3	1	3
ankieta 10	0					1



# Kodowanie pytań zamkniętych z możliwością wyboru kilku wariantów odpowiedzi (**kod geometryczny**) (1)

3. Które kraje europejskie Pan(i) odwiedził(a)?

- |   |     |                                     |       |
|---|-----|-------------------------------------|-------|
| <input type="checkbox"/> Niemcy             | → 1 | <input type="checkbox"/> Grecję     | → 16  |
| <input checked="" type="checkbox"/> Francję | → 2 | <input type="checkbox"/> Hiszpanię  | → 32  |
| <input type="checkbox"/> Węgry              | → 4 | <input type="checkbox"/> Portugalię | → 64  |
| <input checked="" type="checkbox"/> Włochy  | → 8 | <input type="checkbox"/> Inne kraje | → 128 |

Dane: kod arytmetyczny.STA 6v \* 10c

LICZBA	1	2	3	4	5	6
WARTOŚĆ	P_1	P_2	P_3	P_4	P_5	P_6
1			1			
2			16			
3			5			
4			10			
5			3			
6			16			
7			7			
8			8			
9			68			
10			2			

$2+8=10$   
(Francja + Włochy)



## Kodowanie pytań zamkniętych z możliwością wyboru kilku wariantów odpowiedzi (**kod geometryczny**)

Jeżeli respondent wybrał więcej niż jeden wariant, to **łącną odpowiedź** można zakodować za pomocą sumy kodów odpowiadających poszczególnym wariantom odpowiedzi.

Dla celów późniejszej analizy warto od razu sobie zaznaczyć, co oznaczają poszczególne sumy, czyli których kodów stanowią kombinację!!!

Jest to szczególnie ważne przy dużej liczbie możliwych wariantów odpowiedzi, gdyż symbole kodowe narastają w tempie geometrycznym i stają się liczbami wielocyfrowymi.



# Kodowanie pytań zamkniętych z możliwością wyboru kilku wariantów odpowiedzi (**kod geometryczny**) (2)

9. Które z wymienionych czasopism czyta Pan(i) regularnie ?

<input type="checkbox"/>	Angora	→	1	<input type="checkbox"/>	Olivia	→	128
<input checked="" type="checkbox"/>	Auto Motor i Sport	→	2	<input checked="" type="checkbox"/>	Polityka	→	256
<input type="checkbox"/>	Chip	→	4	<input type="checkbox"/>	Przyjaciółka	→	512
<input type="checkbox"/>	Claudia	→	8	<input checked="" type="checkbox"/>	Tele Tydzień	→	1024
<input type="checkbox"/>	Jestem	→	16	<input type="checkbox"/>	Twój Styl	→	2048
<input type="checkbox"/>	Murator	→	32	<input type="checkbox"/>	Życie na Gorąco	→	4096
<input checked="" type="checkbox"/>	National Geographic	→	64	<input type="checkbox"/>	inne czasopismo .....	→	8192

$$2+64+256+1024=1346$$



# Kodowanie pytań zamkniętych z możliwością wyboru kilku wariantów odpowiedzi (**kod binarny (zero-jedynkowy)**) (1)

9. Które z wymienionych czasopism czyta Pan(i) regularnie ?

- |                                     |                     |                                     |                       |
|-------------------------------------|---------------------|-------------------------------------|-----------------------|
| <input type="checkbox"/>            | Angora              | <input type="checkbox"/>            | Olivia                |
| <input checked="" type="checkbox"/> | Auto Motor i Sport  | <input checked="" type="checkbox"/> | Polityka              |
| <input type="checkbox"/>            | Chip                | <input type="checkbox"/>            | Przyjaciółka          |
| <input type="checkbox"/>            | Claudia             | <input checked="" type="checkbox"/> | Tele Tydzień          |
| <input type="checkbox"/>            | Jestem              | <input type="checkbox"/>            | Twój Styl             |
| <input type="checkbox"/>            | Murator             | <input type="checkbox"/>            | Życie na Gorąco       |
| <input checked="" type="checkbox"/> | National Geographic | <input type="checkbox"/>            | inne czasopismo ..... |

**1 = czyta**  
**0 = nie czyta**



Dane: kod arytmetyczny.STA 10v \* 10c

LICZBA	1	2	3	4	5	6	7	8	9	10
WARTO	AMIS	CHIP	CLAUDIA	JESTEM	MURATOR	NAT_GEO	OLIVIA	POLITYKA	PRZYJACI	TELE_TYD
1	1	0	0	0	0	1	0	1	0	1
2										
3										
4										
5										
6										
7										
8										
9										
10										



# Kodowanie pytań zamkniętych z możliwością wyboru kilku wariantów odpowiedzi (**kod binarny (zero-jedynkowy)**) (2)

*Które z wymienionych stacji TV oglądał(a) Pan(i) wczoraj?*

**TVP1**

**TVN**

**Polsat**

**Puls**

<b>Lp.</b>	<b>TVP1</b>	<b>Polsat</b>	<b>TVN</b>	<b>Puls</b>
1	1	0	0	0
2	1	0	1	0
3	1	0	0	0
4	1	1	1	1
5	0	1	1	0
6	1	0	1	0
7	1	1	0	0
8	0	1	0	0
9	0	0	1	1
10	1	1	1	1

# Kodowanie według rodzaju skal

- **skala nominalna** – przypisujemy jedną kolumnę w przypadku skali dysjunkcyjnej (dysjunktywnej), lub tyle kolumn, ile jest pozycji skali w przypadku skali koniunkcyjnej (koniunktywnej).
- **skala porządkowa** - przypisuje się maksymalnie tyle kolumn, ile jest porządkowanych obiektów (cech produktu, marek, firm...). Kolumny odpowiadają rangowanym obiektom (obiekt 1, obiekt 2, obiekt 3,...) lub kolejności rangowej obiektów (miejsce 1, . miejsce 2, miejsce 3,...)
- **skala przedziałowa** - odpowiada jedna kolumna, w której wpisuje się symbole cyfrowe odpowiadające poziomom ocen.
- **skala ilorazowa** - przypisuje się jedną kolumnę - wpisuje się rzeczywiste (niekodowane) wartości cechy.

# Kodowanie odpowiedzi na pytania otwarte

W przypadku pytań otwartych kodowanie ma miejsce dopiero po zebraniu kompletu kwestionariuszy.

Obiektem kodowania nie jest tutaj pojedynczy kwestionariusz ankiety lub wywiadu (jak w przypadku pytań zamkniętych), ale każde kolejne pytanie otwarte z osobna.

Proces kodowania pytań otwartych ma doprowadzić do ich „zamknięcia” przez kodera i traktowania ich w dalszym procesie kodowania jak pytań zamkniętych.

1. **Sporządzenie listy** wszystkich **odpowiedzi** na pytania otwarte;
2. **Uporządkowanie, sklasyfikowanie** odpowiedzi i **sporządzenie listy** ogólnych kategorii powstałych z **jednorodnych grup odpowiedzi** respondentów **według kryterium logiczności**, podobnie jak w przypadku skalowania odpowiedzi na pytania zamknięte;
3. **Przyporządkowanie kodów** poszczególnym **pozycjom listy**.

# Kodowanie odpowiedzi na pytania otwarte

Bardzo  pomocnym narzędziem  przy  kodowaniu  szczególnie  dłuższych  kwestionariuszy jest sporządzenie  **instrukcji kodowej** .

W oparciu o instrukcję kodową  następuje cały  proces   **przyporządkowania symboli kodowych**  poszczególnym  **pytaniom**  kwestionariusza.

Ze względu na to, że  kodowanie pytań zamkniętych  i  pytań otwartych  ma  **inny przebieg** , warto  sporządzić   **osobną instrukcję**  kodową dla  **pytań zamkniętych**  i  osobną  dla  **pytań otwartych** .

**Instrukcja kodowa**  dla  **pytań otwartych**  powinna zawierać dodatkowo  **wskaźniki**   poszczególnych   **kategorii odpowiedzi** .



# Instrukcja kodowa dla pytań otwartych

Nr pytania	Treść pytania	Kategorie odpowiedzi	Wskaźniki odpowiedzi	Symbole kodowe
1.	Jakimi cechami powinien według Pani/Pana charakteryzować się idealny sklep?	a) cechy związane z personelem	"powinna być szybka i miła obsługa"	1
			"sprzedawca musi się znać na tym, co sprzedaje"	
		b) cechy związane z wystrojem sklepu	"powinno być czysto i estetycznie"	2
			"chciałbym, żeby był kolorowy"	
		c) cechy związane z produktem	"jak najwięcej towaru"	4
			"powinien być jak najlepszy towar"	
		d) brak zdania	"nie wiem"	8
"nigdy się nie zastanawiałem"				
e) brak odpowiedzi	-----	16		



# Rodzaje kodów

- kod arytmetyczny
- kod geometryczny
- kod binarny (zero-jedynkowy)
- kodowanie skal złożonych (to Państwu podaruję)

# Kodowanie skal rangowych

Może jeszcze trochę o kodowaniu skali rang.  
Kodowanie to można wykonać na 2 sposoby:

- kiedy zmiennymi będą odpowiedzi
- kiedy zmiennymi będą miejsca, na których wskażą je respondenci

# Kodowanie skal złożonych – RANGOWANIE

*Poniżej wymieniono czynniki brane pod uwagę przy wyborze miejsca zamieszkania. Proszę je uporządkować od, Pana/Pani zdaniem, najważniejszego (przypisując mu numer 1) do najmniej ważnego (przypisując mu numer 4).*

**odległość od pracy**

**renoma okolicy**

**bliskość środków komunikacji publicznej**

**bliskość szkół i przedszkoli**

# Kodowanie skal złożonych – RANGOWANIE (1)

*Poniżej wymieniono czynniki brane pod uwagę przy wyborze miejsca zamieszkania. Proszę je uporządkować od, Pana/Pani zdaniem, najważniejszego (przypisując mu numer 1) do najmniej ważnego (przypisując mu numer 4).*

- 3** odległość od pracy
- 1** renoma okolicy
- 2** bliskość środków komunikacji publicznej
- 4** bliskość szkół i przedszkoli



Lp.	odległość	renoma	komunikacja	szkoły	...
1	<b>3</b>	<b>1</b>	<b>2</b>	<b>4</b>	
2					
3					

## Kodowanie skal złożonych – RANGOWANIE (2)

*Poniżej wymieniono czynniki brane pod uwagę przy wyborze miejsca zamieszkania. Proszę je uporządkować od, Pana/Pani zdaniem, najważniejszego (przypisując mu numer 1) do najmniej ważnego (przypisując mu numer 4).*

- 3** odległość od pracy - 1
- 1** renoma okolicy - 2
- 2** bliskość środków komunikacji publicznej - 3
- 4** bliskość szkół i przedszkoli - 4



Lp.	Miejsce 1	Miejsce 2	Miejsce 3	Miejsce 4	...
1	2	3	1	4	
2					
3					

# Tabulacja (tabelaryzacja) danych

Przykłady, które pojawią się w tej części pochodzą z fragmentu książki:

Bazarnik J., Grabiński T., Kącik E., Mynarski S., Sagan A., **Badania marketingowe. Metody i oprogramowanie komputerowe**, Canadian Consortium of Management Schools – Akademia Ekonomiczna w Krakowie, Warszawa – Kraków 1992



# Tabulacja (tabelaryzacja) danych

**Tabulacja** - zliczanie przypadków należących do różnych kategorii.

Ma ona na celu zestawienie częstości występowania poszczególnych zmiennych i zrelatywizowania w formie procentowej.

**Cel tabulacji:** Organizowanie danych w przejrzystą formę (tabele jednowymiarowe, dwuwymiarowe, wielowymiarowe) i przedstawianie ich w ujęciu procentowym, co ułatwia wizualizację i dalszą analizę statystyczną.

Przeprowadza się ją zarówno dla danych zbieranych zarówno za pomocą tradycyjnych metod (z wykorzystaniem kwestionariuszy papierowych), jak i dla danych gromadzonych przy wykorzystaniu kwestionariuszy elektronicznych.

Tabulacja może przybierać postać **tabulacji prostej** lub **złożonej**.

**Tabulacja prosta** dotyczy zliczania występowania jednej zmiennej.

# Tabulacja danych

Wyróżnia się **3 typy tabel surowych**:

- **tabele jednodzielcze** – przedstawiają one dane o kolejnych zmiennych wyodrębnionych w badaniach;
- **tabele dwudzielcze** – pokazują jednocześnie rozkład dwóch cech;
- **tabele wielodzielcze** – zawarte są w nich jednocześnie rozkłady trzech lub więcej zmiennych.

Ale na przykład w programie **Statistica** tabele dwudzielcze zaliczono do wielodzielczych, w związku z czym są tam **2 typy tabel**:

- **tabele jednodzielcze**
- **tabele wielodzielcze**.

Tabele dwudzielcze i wielodzielcze nazywamy **tabelami krzyżowymi** lub **tabelami kontyngencji**.



# Tabele jednodelcze

<b>Dni tygodnia</b>	<b>Liczba nabywców</b>	<b>Względna liczba nabywców (Procent)</b>
<b>Poniedziałek</b>	3450	23%
<b>Wtorek</b>	2400	16%
<b>Środa</b>	2000	13%
<b>Czwartek</b>	2500	17%
<b>Piątek</b>	4500	30%
<b>Razem</b>	14850	100%



# Tabele dwudzielcze - Podstawowe elementy tabeli

W tabelach wyróżniamy trzy podstawowe elementy:

**Główka** to wiersze opisujące kolumny. Zazwyczaj składa się tylko z jednego wiersza, ale nie musi tak być.

**Boczek** to kolumna opisująca wiersze (czasami jest to więcej niż jedna kolumna). Czasami boczek jest pomijany.

**Rubryki** (komórki) zawierają dane. Łącznie zajmują one największą powierzchnię tabeli.

# Tabele dwudzielcze

**główka**  
↓

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150 8%	1000 7%	1300 9%	500 3%	2000 13%	5950 40%
kobiety	2300 15%	1400 9%	700 5%	2000 13%	2500 17%	8900 60%
suma	3450 23%	2400 16%	2000 13%	2500 17%	4500 30%	14850 <b>100%</b>

**rubryki (komórki)**  
↑

**boczek** →



# Tabele dwudzielcze

W przypadku tabel dwudzielczych **ważnym problemem** przy obliczeniach procentowych jest **wybór kierunku procentowania**.

Możemy mówić o **3 kierunkach procentowania**:

- do całości;
- do wierszy;
- do kolumn.

W większości przypadków, gdy jedna zmienna jest **zmienną niezależną** (przyczyną, zmienną wyjaśniającą lub prognostyczną), a druga zmienna – **zmienną zależną** (skutkiem, zjawiskiem wyjaśnianym lub efektem prognozy), to za podstawę obliczeń można brać **liczebności zmiennej niezależnej**.

Zmienną niezależną zamieszczamy zazwyczaj **w boczku**, a **zmienną zależną w główce**.



## Tabele dwudzielcze (procentowanie do całości)

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150 8%	1000 7%	1300 9%	500 3%	2000 13%	5950 40%
kobiety	2300 15%	1400 9%	700 5%	2000 13%	2500 17%	8900 60%
suma	3450 23%	2400 16%	2000 13%	2500 17%	4500 30%	<b>14850</b> <b>100%</b>

**Procenty odnosimy do sumy całości!**



## Tabele dwudzielcze (procentowanie do wierszy)

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150 19%	1000 17%	1300 22%	500 8%	2000 34%	5950 <b>100%</b>
kobiety	2300 26%	1400 16%	700 8%	2000 22%	2500 28%	8900 <b>100%</b>
suma	3450 23%	2400 16%	2000 13%	2500 17%	4500 30%	14850 <b>100%</b>

### Procenty odnosimy do sum wierszy!

Ile procent mężczyzn dokonuje zakupów w poszczególnych dniach tygodnia i ile procent kobiet dokonuje zakupów w poszczególnych dniach tygodnia?



## Tabele dwudzielcze (procentowanie do kolumn)

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150	1000	1300	500	2000	5950
	33%	42%	65%	20%	44%	40%
kobiety	2300	1400	700	2000	2500	8900
	67%	58%	35%	80%	56%	60%
suma	3450	2400	2001	2500	4500	14850
	100%	100%	100%	100%	100%	100%

### Procenty odnosimy do sum kolumn!

Jaka jest struktura zakupów ze względu na płeć w każdym dniu tygodnia?



# Tabele dwudzielcze (wszystkie 3 sposoby procentowania)

Zauważcie Państwo na kolejnym slajdzie, że pomimo, że w każdej tabeli w tym samym okienku mamy tę samą liczbę osób, to przy różnych kierunkach procentowania, mamy zupełnie różne procenty.

Dlatego musimy się zastanowić, który kierunek procentowania, będzie dla nas najbardziej użyteczny. Czasem możemy zastosować więcej niż jeden.

Musimy też pamiętać, że z różnymi kierunkami procentowania, wiąże się różna interpretacja wyników (proszę sprawdzić na trzech poprzednich slajdach).



# Tabele dwudzielcze (wszystkie 3 sposoby procentowania)

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150	1000	1300	500	2000	5950
	8%	7%	9%	3%	13%	40%
kobiety	2300	1400	700	2000	2500	8900
	15%	9%	5%	13%	17%	60%
suma	3450	2400	2000	2500	4500	14850
	23%	16%	13%	17%	30%	<b>100%</b>

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150	1000	1300	500	2000	5950
	19%	17%	22%	8%	34%	<b>100%</b>
kobiety	2300	1400	700	2000	2500	8900
	26%	16%	8%	22%	28%	<b>100%</b>
suma	3450	2400	2000	2500	4500	14850
	23%	16%	13%	17%	30%	<b>100%</b>

płeć	Zakupy w poszczególne dni tygodnia					
	poniedziałek	wtorek	środa	czwartek	piątek	suma
mężczyźni	1150	1000	1300	500	2000	5950
	33%	42%	65%	20%	44%	40%
kobiety	2300	1400	700	2000	2500	8900
	67%	58%	35%	80%	56%	60%
suma	<b>3450</b>	<b>2400</b>	<b>2001</b>	<b>2500</b>	<b>4500</b>	<b>14850</b>
	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>



# Tabele wielodzielcze – tabela trójdzielcza

płeć	wiek	Zakupy w poszczególne dni tygodnia					suma
		poniedziałek	wtorek	środa	czwartek	piątek	
mężczyźni	do 18 lat	100	50	300	20	200	<b>670</b>
	18-30 lat	500	250	200	80	900	<b>1930</b>
	31-50 lat	400	500	500	100	700	<b>2200</b>
	50 i więcej lat	150	200	300	300	200	<b>1150</b>
kobiety	do 18 lat	300	400	100	500	500	<b>1800</b>
	18-30 lat	500	200	400	500	500	<b>2100</b>
	31-50 lat	1000	400	100	800	1000	<b>3300</b>
	50 i więcej lat	500	400	100	200	500	<b>1700</b>
suma		<b>3450</b>	<b>2400</b>	<b>2000</b>	<b>2500</b>	<b>4500</b>	<b>14850</b>

**Tutaj też możemy procentować we wszystkich kierunkach!**

## **Podsumowując:**

Te trzy etapy (edycja, kodowanie i tabulacja danych) tworzą łańcuch, który zamienia surowe, często chaotyczne informacje w uporządkowany zbiór danych, gotowy do przeprowadzenia głębszych analiz i wyciągnięcia wniosków.

Kolejny wykład poświęcimy na sprawdzanie istnienia istotnych statystycznie zależności pomiędzy zmiennymi za pomocą testu niezależności chi-kwadrat.