Pomiar cech statystycznych - skale pomiarowe

Pomiar

Pomiar statystyczny w najszerszym sensie to przyporządkowywanie obiektom lub zdarzeniom liczb według określonych zasad (Stevens, 1946). W niektórych przypadkach obiektom przyporządkowywane są nazwy kategorii lub symbole a nie liczby.

Po dokonaniu pomiaru możemy powiedzieć, jakie wartości zmiennej (warianty cechy) posiadają badane jednostki.

Pomiaru zmiennej dokonujemy dla każdej badanej jednostki i czynimy to w jednakowy sposób dla wszystkich jednostek (chyba, że celowo projektujemy badanie, z różnymi sposobami pomiaru jakiejś zmiennej)

Przykład: pomiar kilku zmiennych w badaniu European Social Survey (ESS)

W edycji 10 badania ESS (pl: Europejski Sondaż Społeczny), realizowanej w Polsce w 2022 roku (informacje o realizacji badania we wszystkich uczestniczących krajach, dane oraz dokumentacja dostępne są na stronie: https://ess-search.nsd.no/en/study/172ac431-2a06-41df-9dab-c1fd8f3877e7) respondentom zadano między innymi takie pytania:

Skala nominalna: przykłady

zmienna	przykładowe kategorie / poziomy zmiennej
kolor oczu	zielone, niebieskie, piwne, szare
grupa krwi	0, A, B, AB
ulubiony sposób spędzania wolnego czasu	czytanie, aktywność w mediach społecznościowych, nauka, sport, spotkania ze znajomymi
rodzaj umowy o pracę	na okres próbny, na czas określony, na czas nieokreślony
ustrój polityczny kraju	republika, monarchia
marka samochodu	Peugeot, Nissan, Audi, Maserati, BMW

Pamiętaj, że nawet, jeśli w zbiorze danych rodzaje umowy o pracę (czy kategorie jakiejś innej zmiennej nominalnej) zakodujesz liczbami, czyli np. zamiast “na okres próbny” będziesz wpisywać 1, “na czas określony” zamienisz na 2, a “na czas nieokreślony” na 3, zmienna nadal jest **nominalna** i nie można np. obliczać i interpretować średniej wyliczonej z takich danych.

Skala porządkowa: przykłady

A6 Ile mniej więcej czasu spędza Pan na oglądaniu, czytaniu lub słuchaniu wiadomości, albo też innych programów informacyjnych na temat polityki i bieżących wydarzeń? Chodzi o przeciętny dzień. Proszę odpowiedzieć wpisując godziny i minuty. Na przykład: jeśli zajmuje to Panu godzinę i dwadzieścia minut, proszę wpisać:

0 1 2 0

GODZIN MINUT

G G M M

A15 Na ile, Pana zdaniem, system polityczny w Polsce umożliwia ludziom takim jak Pan wpływ na politykę?

1. ☐ W ogóle nie umożliwia
2. ☐ W bardzo małym stopniu umożliwia
3. ☐ W pewnym stopniu
4. ☐ W dużym stopniu
5. ☐ W pełni

A26 Na którą partię lub ugrupowanie głosował Pan w tamtych wyborach do Sejmu?

1. ☐ Koalicja Bezpartyjni i Samorządowcy
2. ☐ Koalicja Obywatelska
3. ☐ Konfederacja Wolność i Niepodległość
4. ☐ Polskie Stronnictwo Ludowe
5. ☐ Prawo i Sprawiedliwość
6. ☐ Sojusz Lewicy Demokratycznej
7. ☐ Na inną partię *(Proszę wpisać)*

dotyczy wyborów z 2019 roku

zmienna	przykładowe kategorie / poziomy zmiennej
wielkość miejscowości zamieszkania	wieś, miasto do 20 tys.mieszkańców, miasto 200 - 500 tys.mieszkańców
częstotliwość chodzenia do teatru	wcale, rzadziej niż raz w roku, kilka razy w miesiącu, częściej niż raz na tydzień
poziom zadowolenia ze studiów	bardzo niezadowolony(a), ani niezadowolony(a) ani zadowolony(a), raczej zadowolony(a)
liczba znajomych na portalu społecznościowym X	poniżej 20, między 21 a 50, między 101 a 200, powyżej 300
klasa energetyczna pralki	A, B, D, F, G
poziom wykształcenia	podstawowe nieukończone i bez wykształcenia szkolnego, podstawowe, średnie, ...

Poziomom zmiennych porządkowych zazwyczaj przypisujemy liczby – rangi (zob. ramkę "Rangowanie – przykłady"). Nie traktujemy jednak tych liczb, jak mających pełne znaczenie matematyczne – służą nam one tylko do ułożenia wartości zmiennej rosnąco/malejąco. Nie ma więc znaczenia, czy poziomom zmiennej przypiszemi rangi: 1, 2, 3, 4, 5, czy np. 0, 18, 19, 200, 236 – oby przypisane liczby odzwierciedlały kolejność wartości – ich uporządkowanie/ranking.

Nie powiemy na przykład, że osoba z wykształceniem wyższym ma o trzy wyższe wykształcenie niż osoba z podstawowym wykształceniem (czy jakimkolwiek innym ;)), albo, że jest dwa razy lepiej wykształcona od osoby z jakimś niższym poziomem wykształcenia. Tak samo dzieje się w przypadku wszystkich zmiennych porządkowych. **Nie liczymy dla nich średniej**, ani miar na niej opartych (np. odchylenia standardowego, czy współczynnika asymetrii).

A62 Ile jest takich osób, z którymi może Pan otwarcie porozmawiać o wszystkich sprawach osobistych i związanych z pracą?

1. ☐ Nie ma takich osób
2. ☐ 1
3. ☐ 2
4. ☐ 3
5. ☐ 4-6
6. ☐ 7-9
7. ☐ 10 lub więcej osób

D5-9 W jakim stopniu, Pana zdaniem, komunikacja internetowa i mobilna…

D5 …sprawia, że ludzie czują się sobie bliżsi?

W ogóle nie — W pełni

0 1 2 3 4 5 6 7 8 9 10

C22 Czy w swoim głównym miejscu pracy kieruje/ kierował Pan pracą innych osób?

1. ☐ Tak ⟶ **Proszę odpowiedzieć na C23 poniżej**
2. ☐ Nie ⟶ **Proszę przejść do C24-25 poniżej**

C23 Pracą ilu osób Pan kieruje/ kierował?

Co różni powyższe pytania?

Oczywiście temat, ale poza tym?

Na pewno różna jest liczba możliwych odpowiedzi, czyli uzyskamy zmienne o różnej liczba kategorii / poziomów. W pytaniu C22 mamy np. tylko dwie opcje: tak i nie, w pytaniu A26 wybierano z siedmiu możliwości, w D5 z jedenastu a np. w



Skala interwałowa: przykłady

zmienna	przykładowe kategorie / poziomy zmiennej
rok założenia firmy	1897, 1990, 2013, 2021
poziom zadowolenia ze studiów w skali 0 - 10	0, 3, 7, 8, 10
wynik pomiaru IQ skalą Wechslera	92 , 100, 110, 135
temperatura powietrza w °C	2°C, -12°C, 27°C, 33°C
ocena jakości produktu w skali od 1 do 7	1, 2, 4, 5, 7
godzina rozpoczynania pracy	7:30, 8:00, 9:30, 11:00

Zauważ, że w przypadku żadnej ze zmiennych wymienionych w powyższej tabeli zero nie oznacza absolutnego braku mierzonego zjawiska. To mamy na myśli mówiąc, że w pomiarze interwałowym zero jest **zerem względnym***, a nie absolutnym – ktoś arbitralnie ustalił punkt początkowy skali. Tak, jak w przypadku stopni Celsjusza, gdzie zero jest temperaturą jednej ze zmian stanu skupienia wody. 0° Celsjusza nie oznacza braku temperatury. Ponadto zero “znika” (dostajemy w pomiarze inną wartość liczbową niż 0), jeśli zmienimy termometr na taki, który pokazuje wartości w skali Fahrenheita (wtedy zamiast 0°

pytaniu A6 możliwości jest bardzo wiele – każdy może wpisać w zasadzie dowolną liczbę (nie przekraczającą liczby godzin i minut, jakie ma doba).

Ok, ale to nie wszystko.

Wyobraź sobie, że te pytania zadajesz wszystkim pracującym studentom z Twojego kierunku. Chcesz następnie podsumować wyniki badania. **Co można zrobić z danymi po uzyskaniu odpowiedzi na każde z pytań**?

Mając informacje o **preferencjach partyjnych** swoich respondentów (A26), możesz sprawdzić jaki **procent** osób badanych zadeklarował, że głosował na daną partię. Wskazać **najmniej i najbardziej "popularną"** partię w badanej grupie. Możesz też określić na ile **różnorodne** były odpowiedzi – czy większość respondentów głosowała na którąś z dwóch największych partii, czy może głosy rozkładały się równomiernie. Ale na tym koniec. **Nie policzymy tu np. średniej ani mediany.** A o dwóch badanych osobach można powiedzieć tylko, czy udzieliły takiej samej odpowiedzi, czy różnych.

Inaczej jest np. w przypadku pytania A6. Jeśli przeliczymy uzyskane wyniki na minuty (czyli będziemy mieć w zbiorze np. 90 minut zamiast 1h 30 minut), możemy nie tylko powiedzieć, jaka odpowiedź padała **najczęściej**, a jaka **najrzadziej**, ale policzyć także ile czasu **średnio** badani spędzali dziennie (według swoich deklaracji) na zapoznawaniu się z wiadomościami na temat polityki i bieżących wydarzeń, obliczyć **odchylenie standardowe** tej zmiennej w zbiorze, wskazać **medianę i kwartyle,** odpowiedzieć na pytanie, jaki % osób poświęcał na

zobaczymy 32°). Uzyskanie w pomiarze zera zależy więc od zastosowanej skali. Co więcej nie każda skala w pomiarze interwałowym musi zawierać 0.

(*Trudno jest nam też sobie wyobrazić na przykład, czym miałby być całkowity i absolutny brak inteligencji...)

Dla zmiennych interwałowych **możemy liczyć średnią**. Można obliczyć przeciętny poziom zadowolenia ze studiów w jakiejś grupie (jeśli był on mierzony np. tak, jak wskazano w tabeli) lub przeciętną temperaturę powietrza zarejestrowaną np. w danej stacji meteorologicznej w południe w kolejnych dniach lipca. W niektórych przypadkach jednak – jak np. dla zmiennej "rok" (założenia firmy, urodzenia...), choć teoretycznie średnią policzyć i zinterpretować by się dało, w praktyce raczej się tego nie robi. Do obliczeń matematycznych i formułowania na ich podstawie konkluzji użylibyśmy tu najpewniej zmiennej ilorazowej wyrażającej liczbę lat (obecności na rynku, czy wieku osoby).

Skala ilorazowa: przykłady

zapoznawanie się z wiadomościami więcej niż np. 2h dziennie, sprawdzić, czy w zbiorze dominują względnie niskie, zbliżone do średniej, czy wysokie wartości określające liczbę minut (a może żadne z nich). Dla dowolnych dwóch badanych osób można powiedzieć nie tylko, czy podały taką samą liczbę, czy różne, ale, jeśli odpowiedzi były różne, określić także, że jedna osoba poświęca więcej czasu niż druga, a ponadto wskazać o ile minut więcej i obliczyć ile razy więcej (jeśli osoba A odpowiedziała, że 30 minut, a osoba B, że 90, oznacza to, że osoba B poświęcała na zapoznawanie się z wiadomościami trzy razy więcej czasu niż osoba A).

Wszystkie te kwestie związane są ze skalą pomiarową zmiennej. Z punktu widzenia analizy statystycznej kluczowe jest to, że wymienione wyżej zmienne różnią się poziomem pomiaru (skalą pomiarową).

Skale pomiarowe

Informacje na temat interesujących nas zmiennych gromadzić można z różną szczegółowością. Zakres informacji o relacjach między badanymi obiektami, jakie uzyskujemy w wyniku pomiaru określany jest przez poziom pomiaru inaczej skalę pomiarową i przekłada się na możliwe do zastosowania metody analizy danych (dlatego to takie ważne zagdanienie!).

Skale pomiarowe w statystyce dzielimy na typy, które można uporządkować według mocy (poniżej od najsłabszej do najsilniejszej). Moc skali związana jest z możliwościami

zmienna	przykładowe kategorie / poziomy zmiennej
czas poświęcany zazwyczaj dziennie na czytanie książek	0, 15, 25, 45, 90, 180 (minut)
metraż mieszkania	23.5, 37.0, 56.2, 7.06, 92.8 (m^2)
liczba podejść do egzaminu praktycznego na prawo jazdy	0, 2, 7, 12
wysokość miesięcznych wydatków na...	0, 10, 23, 40, 85, 112, 160 (zł)
procentowy wynik z egzaminu z przedmiotu A	0, 33, 52, 68, 75, 85, 94, 100 (%)
ilość rocznie zużytej wody (na mieszkańca gosp.domowego)	15.3, 22.8, 33.7, 35.3, 38.9, 41.3 (m^3)

Zmienna dychotomiczna, zmienna binarna, dummy coding

Zmienne dychotomiczne, to takie, które przyjmują tylko dwie wartości (np. posiadane wykształcenie: średnie / wyższe (jeśli inne osoby nie były badane), rok studiów uzupełniających magisterskich, na którym jest student: 1 / 2, posiadanie prawa jazdy kategorii B: nie / tak, wynik poprawnie wykonanego testu na covid: pozytywny / negatywny, uzyskiwanie dochodów wyższych od średniej krajowej: nie / tak... itd.*).

Dla uproszczenia, podczas zajęć ze statystyki to, czy zmienna jest dychotomiczna, czy nie będziemy określali dla naszych danych, czyli według liczby wartości zmiennej, które mamy **w zbiorze**.

Do oznaczenia liczbami wartości takich zmiennych w zbiorze, czyli liczbowego ich **zakodowania**, bardzo często używa się

dokonywania przekształceń na skali, co z kolei wynika bezpośrednio z ilości informacji, którą skala zawiera*.

nominalna → porządkowa → interwałowa → ilorazowa
/przedziałowa/ /stosunkowa/

Możemy spotykać się w literaturze z różnymi sformułowaniami:

- miary/mierniki nominalne, porządkowe...,
- pomiar nominalny, porządkowy...,
- skala nominalna, porządkowa... ,
- zmienne nominalne, porządkowe...

O skali pomiarowej będziemy mówić dopiero wtedy, gdy zmienna przyjmować będzie **więcej wartości niż dwie**. Jeśli jakaś analizowana przez Ciebie cecha ma tylko dwa warianty/ poziomy (kierunek ruchu na ulicach kraju: prawo-/lewostronny, płeć osoby: kobieta/mężczyzna, wynik losowania: orzeł/reszka, wynik egzaminu: zdał(a)/nie zdał(a), posiadanie wyższego wykształcenia: nie/tak, liczba pełnych ukończonych lat życia przez dziecko urodzone 2019 roku, gdy badanie realizujemy w pewnym momencie roku 2023: 3/4 itd...), określ ją po prostu jako **zmienną dychotomiczną**.**

* Więcej na temat mocy skali możesz przeczytać np. w podręczniku "Podstawy statystyki dla socjologów. Tom 1. Opis statystyczny" autorstwa Grzegorza Lissowskiego, Jacka Hamana i Mikołaja Jasińskiego (Warszawa: Scholar 2011)

** Na ten temat więcej znajdziesz w "kafelku" "Zmienna dychotomiczna, zmienna binarna, *dummy coding*"

wartości 0 i 1. To tak zwane kodowanie zerojedynkowe (*dummy coding*)

- **0** interpretowane jest wówczas jako **nie** (czyli: nie ma / nie posiada / nie spełnia danego warunku itd.) lub, posługując się językiem logiki, **fałsz (FALSE);**
- **1** natomiast oznacza **tak** (czyli: ma / posiada / spełnia dany warunek itd.) lub: **prawda (TRUE)**.

Taki zabieg ma największy sens i stanowi ułatwienie w interpretacji wyników analiz, jeśli kodowaniu zerojedynkowemu towarzyszy **odpowiednia nazwa zmiennej**. Warto też od razu podkreślić, że **każdą zmienną** można przekodować na zmienną lub zmienne binarne. (Zobacz "kafelek": "Kodowanie zerojedynkowe – przykłady")

* przykłady zmiennych dychotomicznych są też w "kafelku": "Skale pomiarowe"



Kodowanie zerojedynkowe - przykłady

Problemowi **odpowiedniej nazwy** zmiennej zerojedynkowej, przyjrzyjmy się na przykładzie zmiennej "płeć". Gdy chcemy zakodować ją jako *dummy,* kiepskim pomysłem jest

Skala nominalna

Umożliwia tylko klasyfikację/kategoryzację jednostek ze względu na posiadanie przez nie określonych wariantów cechy.

Nie daje możliwości uporządkowania wariantów cechy ani nie posiada precyzyjnej jednostki pomiaru.

Jeśli zmienna zmierzona została na skali nomialnej, dla dwóch jednostek możliwe jest jedynie określenie, czy należą do tej samej klasy/kategorii (czyli, czy mają jednakowy wariant danej cechy) czy nie. Jedyne symbole, których można użyć do opisania relacji między obiektami w tym przypadku to: = i ≠

Przykładowo, jeśli spytamy grupę osób o miasto urodzenia, dla każdej pary badanych możemy powiedzieć, czy te dwie osoby urodziły się w tym samym mieście, czy w różnych miastach. Nie można np. uporządkować zbioru rosnąco ani malejąco wg wartości zmiennej miasto urodzenia (można byłoby wg wielkości miejsca urodzenia, albo np. jego szerokości geograficznej, ale to już inne zmienne!), ani powiedzieć, że osoba A ma wyższą lub niższą wartość tej zmiennej od osoby B.



pozostawienie nazwy “płeć”, ponieważ konieczne jest później pamiętanie, którą płeć oznaczyliśmy jako 0, a którą jako 1.

Jeśli zaś nazwiemy zmienną *dummy* “kobieta” lub np. “kobieta(0/1)” wiadomo będzie, że wartość 1 przypisana została osobom, które np. zaznaczyły w kwestionariuszu, że są kobietami, a 0 wszystkim innym – czyli “niekobietom”.



Co tyczy się zaś możliwości **zakodowania każdej zmiennej jako *dummy***, przyjrzyjmy się przykładom zmiennych: “wiek” oraz “wielkość miejscowości zamieszkania”, których pomiaru dokonano w pewnej grupie uczestników forum dyskusyjnego na jednym z internetowych portali. Pierwszą ze zmiennych z ilorazowej przekodowano na binarną: “pełnoletni(a)”, z drugiej utworzone zostały dwie zmienne binarne: “mieszka na wsi” i “mieszka w dużym mieście (100 tys. mieszkańców i więcej)”.

nr obs.	płeć	kobieta	wiek	pełnoletni(a)	wielkość miejscowości zamieszkania	mieszka na wsi	mieszka w dużym mieście (100K miesz.+)
1	mężczyzna	0	22	1	Miasto powyżej 500 tys. mieszk.	0	1
2	kobieta	1	15	0	Miasto 20-100 tys. mieszk.	0	0
3	mężczyzna	0	18	1	Wieś	1	0
4	mężczyzna	0	17	0	Miasto 20-100 tys. mieszk.	0	0
5	kobieta	1	20	1	Miasto 200-500 tys. mieszk.	0	1
6	kobieta	1	22	1	Miasto powyżej 500 tys. mieszk.	0	1
7	kobieta	1	19	1	Wieś	1	0
8	mężczyzna	0	18	1	Miasto poniżej 20 tys. mieszk.	0	0
9	kobieta	1	16	0	Miasto 200-500 tys. mieszk.	0	1
10	mężczyzna	0	16	0	Miasto 100-200 tys. mieszk.	0	1

Skala porządkowa

Umożliwia klasyfikację i porządkowanie jednostek rosnąco/ malejąco według wartości zmiennej. Brak jednostki pomiaru.

Jeśli zmienna została zmierzona na skali porządkowej, dla dwóch jednostek, u których dokonano pomiaru możliwe jest określenie, czy mają taki sam poziom zmiennej / wariant cechy, czy różne, a jeśli się różnią można także powiedzieć, czy jednostka A ma wyższy, czy niższy poziom zmiennej niż jednostka B.

Relacje między obiektami wyraża się więc stosując symbole:
$= lub \neq$ oraz $> lub <$

Gdy chce się, by zmienna porządkowa była jako taka właśnie traktowana w analizie statystycznej (np. wykonanej w programie komputerowym), zastosować trzeba tzw. rangowanie – przyporządkowanie każdej wartości zmiennej liczby wyrażającej jej miejsce w uporządkowanym szeregu. Rangowanie odzwierciedla "nie tylko równość i różność obiektów, lecz również możliwość ich uporządkowania (przedstawienia rankingu) ze względu na intensywność mierzonej własności." (Lissowski, Haman, Jasiński, 2011, s. 36). (więcej na ten temat przeczytasz w ramce "Rangowanie – przykłady", znajdującej się niżej)

Skala interwałowa (przedziałowa)

Umożliwia klasyfikację i porządkowanie jednostek według wartości zmiennej (czyli tak, jak skala porządkowa), ale dodatkowo pozwala na oraz porównywanie wielkości różnic

Zauważ, że dla zmiennej zerojedynkowej można policzyć np. średnią: w przypadku zmiennej kobieta otrzymamy średnią 0,5, dla "mieszka na wsi": 0,2. Co oznaczają te wyniki / co mówią o analizowanym zbiorze?

Obecność liczb nie definiuje skali pomiaru

W poniższym zestawieniu (jest to fragment przykładowego zbioru danych) są tylko zmienne nominalne i porządkowe. Pomimo, że widoczne w danych są liczby, policzenie np. średniej arytmetycznej dla którejkolwiek z tych zmiennych byłoby błędne (!) Liczby pełnią tu jedynie fukncję kodów upraszczających zapis kategorii (!). "Województwo zamieszkania" jest nadal zmienną **nominalną** a "wielkość miejscowości zamieszkania": **porządkową...** itd

województwo zamieszkania	wielkość miejscowości zamieszkania	płeć	rodzaj umowy o pracę	poziom wykształcenia
3	6	2	3	4
11	4	2	3	2
4	5	1	2	3
4	2	2	1	1
1	6	1	2	1
6	2	1	2	3
7	1	1	3	2
3	1	2	2	3
2	6	1	1	3

Podobnie nie liczylibyśmy średniej ani innych miar statystycznych dla zmiennej "numer bloku", czy "PESEL".

pomiędzy jednostkami. Jednostka pomiaru; Odstępy między kolejnymi wartościami na skali są takie same, co oznacza, że na zmiennych interwałowych możemy wykonywać **działania matematyczne**, takie jak dodawanie, odejmowanie lub obliczanie średniej arytmetycznnej.

Początek skali stanowi tzw. zero względne (jest to po prostu arbitralnie ustalony początek skali).

Jeśli zmienna została zmierzona na skali interwałowej, dla dwóch jednostek u których dokonano pomiaru możliwe jest określenie, czy mają taki sam poziom zmiennej / wariant cechy, czy różne, a jeśli się różnią można także powiedzieć, czy jednostka A ma wyższy, czy niższy poziom zmiennej niż jednostka B oraz policzyć **o ile** wyższy/niższy poziom zmiennej ma jednostka A w porównaniu do jednostki B.

Relacje między obiektami wyraża się więc stosując symbole: $= lub \neq$ oraz $> lub <$. Choć w interpretacji różnic w poziomach zmiennej interwałowej między jednostkami **nie** możemy używać sensownie zwrotów wyrażających stosunki wartości (np. dwa razy więcej, 3,5 raza mniej); interpretować możemy natomiast **stosunki różnic**:

$$\frac{x_1 - x_2}{x_3 - x_4}$$

Nie powiemy np. że rok 2000 był dwa razy później niż rok 1000, albo że to dwa razy więcej czasu (to byłaby – bardzo nieudana - próba interpretacji stosunku dwóch wartości zmiennej). Możemy zaś powiedzieć, że między rokiem 2010 a 2022 upłynęło dwa

Wyliczona z uśredniania liczb wartość nie miałaby w tym przypadku żadnej sensownej interpretacji.

Dlatego ZAWSZE trzeba sie przyjrzeć zmiennym i ich wariantom/ wartościom/poziomom, aby zorientować się co dokładnie kryje sie za widocznymi w zbiorze danych kodami.

Powiązania między skalami pomiarowymi

“Przechodzenie” między różnymi poziomami pomiaru jest możliwe jedynie przy przejściu z silniejszej skali pomiarowej na słabszą (nigdy odwrotnie!).

Jeśli np. dokonamy pomiaru miesięcznych zarobków netto pewnej grupy osób, prosząc badanych o podawanie dokładnych kwot (np. w zł, zaokrąglonych do wartości całkowitych) pomiar ten określimy jako ilorazowy. Będzie można wówczas policzyć na przykład średnie zarobki w badanej grupie (wraz z odchyleniem standardowym rozkładu, wyrażającym rozrzut wyników pomiaru wokół średniej) itd.

Pogrupowanie już zgromadzonych danych w przedziały (np. poniżej 3 tys. zł, 3 – 4,5 tys. zł itd...) sprawi, że zmienna, którą otrzymamy będzie na skali porządkowej. (Tu średniej ani odchylenia standardowego już nie policzymy.) Zakodowanie wyników zerojedynkowo, pozwoli nam z kolei uzyskać zmienną *dummy* np.: “zarobki powyżej 6 tys. zł netto miesięcznie”, które przyjmie wartość 1 dla wszystkich osób, które takie zarobki

razy więcej czasu, niż między rokiem 1990 a 1996 (to interpretacja stosunku różnic, o której mowa wyżej).



Skala ilorazowa (stosunkowa)

Umożliwia klasyfikację i porządkowanie jednostek oraz porównywanie wielkości różnic (czyli wszystko to, co skala interwałowa), a ponadto pozwala na interpretację i porównywanie **stosunków wartości zmiennej pomiędzy jednostkami.** Naturalna jednostka miary z **zerem absolutnym** (zero oznacza *brak*). Możliwe wykonywanie **wszystkich działań matematycznych.**

Jeśli zmienna została zmierzona na skali ilorazowej, dla dwóch jednostek u których dokonano pomiaru możliwe jest określenie:

- czy mają taki sam poziom (wartość) zmiennej, czy różne...
- ...jeśli się różnią można także powiedzieć, czy jednostka A ma wyższy, czy niższy poziom zmiennej niż jednostka B...
- policzyć **o ile** wyższy/niższy poziom zmiennej ma jednostka A w porównaniu do jednostki B

zadeklarowały i 0 w przypadku tych, którzy podali kwoty 6 tys. zł lub niższe.

Jeśli w badaniu poprosimy jedynie o określenie, czy ktoś uzyskuje zarobki powyżej 6 tys. zł miesięcznie, czy nie lub o wskazanie jednego z kilku zdefiniowanych przez nas przedziałów zarobków, nie "wyciągniemy" z tych danych informacji o dokładnej wartości zarobków każdej z osób (czyli ze zmiennej porządkowej lub dychotomicznej nie możemy "przejść" na zmienną ilorazową)

Rangowanie - przykłady (zmienna porządkowa i ilorazowa)

Wyobraźmy sobie, że grupie 8 osób zadano dwa pytania z ESS:

A15: Na ile, Pana zdaniem, system polityczny w Polsce umożliwia ludziom takim, jak Pan wpływ na politykę

A6: Ile mniej więcej czasu spędza Pan na oglądaniu, czytaniu lub słuchaniu wiadomości, albo też innych programów informacyjnych na temat polityki i bieżących wydarzeń?

(dane widoczne są na screenie poniżej).

Przyjrzyjmy się, jak mogłoby wyglądać rangowanie w tych przypadkach.

Możemy:

oraz

- policzyć **ile razy** wyższy/niższy jest poziom zmiennej u jednostki A w porównaniu do jednostki B. Te ilorazy (stosunki) mogą być interpretowane w kategoriach proporcji nasilenia mierzonej własności, na co nie pozwalała skala interwałowa (patrz niżej).

Widząc w na mapie pogody w mieście M1 12°C, a w mieście M2 36°C (skala interwałowa) nie powiemy, że w tym drugim było trzy razy cieplej niż w pierwszym. Jeśli natomiast jedna miejscowość ma powierzchnię 2,34 km² (skala ilorazowa), a inna 4,68 km² możemy powiedzieć, że ta druga jest dwa razy większa (ma dwa razy większą powierzchnię) niż pierwsza.

Warunki prawidłowego określania kategorii/ poziomów zmiennej

Wartości (poziomy) zmiennej na każdej ze skal muszą być względem siebie:

1. ROZŁĄCZNE
   każdą obserwację można przypisać do jednej i tylko jednej wartości zmiennej

oraz

2. WYCZERPUJĄCE
   musimy móc przyporządkować każdej obserwacji

- rangować same wartości/poziomy zmiennej (uczynić to można nawet jeszcze przed badaniem); mamy wówczas tyle rang ile poziomów ma zmienna (jak np. w przykładzie dla zmiennej A15, rangi to: 1,2,3,4,5)

- rangować obserwacje według wartości zmiennej; najwyższa ranga odpowiada wtedy numerowi ostatniej obseracji w zbiorze (jak np. w przykładzie dla zmiennej A6), a przypisywane rangi odzwierciedlają pozycje (numery wierszy) zajmowane przez jednostki po uporządkowaniu danych wg wartości tej zmiennej*. To częstsze podejście w przypadku zmiennych ilorazowych.

odpowiednią wartość zmiennej

Źródła

Icons made by Freepik from www.flaticon.com:

https://www.flaticon.com/authors/freepik

osoba badana	A15	ranga poziomu zmiennej A15	A6 (w minutach)	ranga obserwacji wg zmiennej A6
1	W dużym stopniu	4	30	4
2	W ogóle nie umożliwia	1	15	2,5
3	W pełni	5	50	5
4	W pewnym stopniu	3	180	8
5	W ogóle nie umożliwia	1	5	1
6	W dużym stopniu	4	60	6
7	W bardzo małym stopniu umożliwia	2	15	2,5
8	W ogóle nie umożliwia	1	90	7

W przypadku zmiennej A15 rangi odzwierciadlają jedynie porządek wartości zmiennej i pochodzą stąd:

A15 Na ile, Pana zdaniem, system polityczny w Polsce umożliwia ludziom takim jak Pan wpływ na politykę?

1. ☐ W ogóle nie umożliwia
2. ☐ W bardzo małym stopniu umożliwia
3. ☐ W pewnym stopniu
4. ☐ W dużym stopniu
5. ☐ W pełni

Dla zmiennej A6 porangowano obserwacje. Najpierw dane zostały uporządkowane wg wartości zmiennej (liczba minut...) rosnąco:
5 **15** **15** 30 50 60 90 180
A później każdej wartości przypisano jej miejsce w uporządkowanym szeregu z uwzględnieniem pozycji zajmowanej przez wynik(i) w danych:
Rangi:
1 **2,5** **2,5** 4 5 6 7 8

(liczba 15 zajmuje miejsce 2 i 3.
Z uśrednienia tych liczb dostajemy 2,5)

* osoba, która w pytaniu A6 podała odpowiedź 5 minut, ma najniższą wartość zmiennej, zajmowałaby pierwszą pozycję, gdybyśmy uporządkowali dane rosnąco, "otrzymuje" więc rangę 1. Dwie osoby, które podały odpowiedź 15 minut, po uporządkowaniu danych znajdowałyby się w linijkach 2 i 3. **Udzieliły tej samej odpowiedzi, więc rangę muszą mieć tę samą**. Najczęściej w tym przypadku uśredniamy wartości, stąd ranga 2,5.