

# **Metodologia badań naukowych w medycynie**

**Dariusz Radomski  
Antoni Grzanka**

## Spis treści

1. Wstęp.....	3
2. Poznanie naukowe w medycynie od starożytności do współczesności .....	5
2.1 Teoria poznania a nauka.....	5
2.2 Poznanie naukowe w medycynie .....	7
3. Obserwacja jako źródło danych pomiarowych .....	12
3.1 Obserwacja jako źródło informacji .....	12
3.2 Pomiar jako metoda obiektywizacji obserwacji.....	17
3.3 Pomiary wykorzystywane w medycynie.....	20
3.4 Obserwacja i pomiar w praktyce klinicznej .....	21
4. Błąd pomiarowy i niepewność wyników w badaniach naukowych i praktyce klinicznej ..	23
4.1 Definicja błędu pomiarowego .....	23
4.2 Źródła błędów pomiarowych .....	24
4.3 Klasyfikacja błędów pomiarowych.....	25
4.4 Zmienność obserwacji a błąd pomiarowy .....	27
4.5 Błędy pomiarowe a niepewność wyników .....	28
4.6 Analiza rzetelności wyników obserwacji i pomiarów.....	28
5. Zdrowie i choroba jako przedmiot poznania naukowego .....	32
5.1 Pojęcie modelu procesu chorobowego .....	32
5.2 Hierarchiczny opis procesu chorobowego .....	37
5.3 Tworzenie modelu badanego procesu .....	38
5.4 Klasyfikacja modeli.....	39
6. Podstawy wnioskowania w naukach medycznych.....	42
6.1 Rodzaje wnioskowania naukowego .....	42
6.2 Wnioskowanie statystyczne .....	43
6.3 Wybór próby z populacji generalnej .....	44
6.4 Estymacja parametrów .....	45
6.5 Testowanie hipotez.....	46
6.6 Ocena zależności między obserwacjami .....	47
6.7 Pułapki wnioskowania statystycznego .....	48

## 1. WSTĘP

Rozwój nauk przyrodniczych, społecznych i technicznych przyczynia się do wydłużania się długości życia osób oraz dostarcza wiedzy i możliwości dla jego wysokiej jakości. Dążenie cywilizowanych społeczeństw do poprawy jakości życia od chwili poczęcia (rozwój medycyny prenatalnej) aż do chwili śmierci (medycyna geriatryczna i paliatywna) wymusza oczekiwanie wzrostu jakości usług medycznych oraz usług pielęgnacyjnych. Jednym z czynników decydujących o poziomie udzielanych świadczeń zdrowotnych jest wykorzystanie współczesnej wiedzy oraz metodologii poznania naukowego w codziennej praktyce lekarza, pielęgniarki, położnej lub pedagoga profilaktyki zdrowotnej.

Przedstawiony w skrypcie materiał ma na celu przybliżyć Czytelnikom metodologię poznania naukowego stosowaną w medycynie. Elementem łączącym wszystkie rozdziały skryptu jest pojęcie *modelu*, który jest podstawowym narzędziem naukowego opisu rzeczywistości.

Aby lepiej zrozumieć wzajemną koegzystencję sztuki i nauki w medycynie, w rozdziale pierwszym przedstawiono historyczną analizę roli poznania naukowego w tej dziedzinie nauki. Rozdział drugi i trzeci poświęcono omówieniu znaczenia jakości obserwacji dokonywanych w medycynie dla poprawności budowanych modeli teorii naukowych. Rozdział czwarty wprowadza formalną definicję pojęcia modelu, omawia procedurę tworzenia modeli w medycynie, a także przedstawia ich klasyfikację. W rozdziale piątym przedstawiono podstawy wnioskowania (naukowego), które służy właściwej interpretacji wyników uzyskiwanych podczas naukowej analizy badanego procesu biologicznego. W kolejnych dwóch rozdziałach zaprezentowano podstawy teoretyczne przydatne przy wyborze odpowiedniego schematu realizowanego badania naukowego, który determinuje możliwe do uzyskania obserwacje. W rozdziale dziewiątym zamieszczono wiedzę pomagającą w zaprojektowaniu planu badawczego, z uwzględnieniem analizy potrzeb, analizy zasobów sprzętowych i ludzkich. Końcowym efektem poznania naukowego jest publikacja naukowa, pozwalająca na rozpowszechnienie nowych faktów. Podstawowe zasady korzystania i przygotowywania piśmiennictwa naukowego przedstawiono w rozdziale dziesiątym. Rozdział jedenasty pozwoli Czytelnikowi zapoznać się z obszarami medycyny, w których ma zastosowanie modelowanie matematyczne. Celem rozdziału dwunastego jest wykazanie, że obecnie metodologia poznania naukowego odgrywa także istotną rolę w codziennej praktyce lekarskiej.

Skrypt „Metodologia badań naukowych w medycynie” został napisany jako materiał dydaktyczny do przedmiotu o takiej samej nazwie, który prowadzony jest na Wydziale Nauk o Zdrowiu Akademii Medycznej w Warszawie. Może on być również pomocny dla słuchaczy studiów doktoranckich w opracowywaniu i realizacji własnych projektów badawczych.

Autorzy pragną podziękować wszystkim studentom za wnikliwą recenzję maszynopisu skryptu, która znacznie ułatwiła prace redakcyjne.

*Autorzy*

## 2. POZNANIE NAUKOWE W MEDYCYNIE OD STAROŻYTNOŚCI DO WSPÓŁCZESNOŚCI

### 2.1 Teoria poznania a nauka

Działalność naukowa jest jedną z form twórczej aktywności człowieka, mającą na celu poznanie i opanowanie rzeczywistości. Zgodnie z pierwszym zdaniem „Metafizyki” Arystotelesa, „Ludzie z natury dążą do poznania”. Myśl ta przetrwała do czasów współczesnych. W XX wieku, w opracowanej przez Masłowa hierarchii głównych potrzeb człowieka, znalazła się również potrzeba poznania.

Poznanie otaczającego nas świata może odbywać się na wiele sposobów, jednak nie każdy z nich jest akceptowany przez działalność naukową. Teorią poznania zajmują się dział filozofii zwany *epistemologią*. Rozważania epistemologiczne szukają odpowiedzi na pytanie, jaki jest najlepszy, najbardziej obiektywny sposób poznania rzeczywistości. Z epistemologią silnie związana jest *ontologia*, która jest teorią bytu. Głównym problemem ontologicznym jest poszukiwanie odpowiedzi na pytania o strukturę i naturę rzeczywistości. Zatem rozważania epistemologiczne i ontologiczne stanowią bazę filozoficzną działalności naukowej.

Metody poznawania otaczającego świata zmieniają się zarówno w trakcie rozwoju osobniczego człowieka, jak również rozwoju całej cywilizacji. W początkowym okresie życia poznanie ma wyłącznie charakter empiryczny i dokonuje się za pomocą zmysłów. Wraz z rozwojem funkcji poznawczych i kognitywnych, poznanie empiryczne jest uzupełnianie przez poznanie rozumowe.

Pozorny spór między poznaniem *empirycznym* a poznaniem *rozumowym* towarzyszy historii ludzkości od czasów starożytnych po współczesność. Jest on także podstawą różnych koncepcji epistemologicznych nauki.

Pierwszym myślicielem, który rozważał zagadnienia epistemologiczne był Platon. W stworzonej przez niego teorii idei, świat rzeczywisty to pewne idee, niematerialne i ponadczasowe ‘prawzory’ postrzeganych zmysłami obiektów. Według Platona, poznanie świata odbywa się na gruncie pozazmysłowym, idee są wyobrażeniami poznawanych obiektów. Platońskie idee można utożsamiać z tworem ludzkiego umysłu, całkowicie oderwanymi od rzeczywistych reprezentantów danej idei. Tylko niezmienny porządek idei może tworzyć efekt ludzkiego poznania rzeczywistości, zwaną *wiedzą*. W ten sposób Platon zapewnia postulat obiektywnego poznania naukowego, które opiera się na rozumie (wyobrażeniu, idei).

Polemiczną do Platona koncepcję epistemologiczną przedstawił Arystoteles, który uważał, że podstawą poznania i rozumienia świata jest postrzeganie przez zmysły realnych obiektów. Obiekty te są zbudowane z dwóch substancji – materii i formy. Wszelkie poznanie pozazmysłowe tworzy fikcję nie mającą nic wspólnego z nauką.

Obie powyższe koncepcje epistemologiczne zakładały, że poznanie jest kompletne tzn. dostarcza pełną, niezakłóconą subiektywizmem wiedzę o przedmiocie poznania. Przekonanie to zostało mocno osłabione przez koncepcje wysuwane przez Sceptyków i Sofistów, według których należało się wstrzymywać od wszelkich sądów o otaczającej rzeczywistości.

Zwątpienie w możliwość rozumnego poznania rzeczywistości było bardzo silnie akcentowane w Średniowieczu, stanowiąc dobre podłoże filozoficzne i społeczne do ekspansji chrześcijaństwa na całą Europę. Ujawniający się w chrześcijaństwie skrajny antyintelektualizm oraz fideistyczny sposób postrzegania świata zatrzymał rozwój epistemologii na całe Średniowiecze. Poznanie świata był wówczas utożsamiane z poznaniem Boga, a za prawdę uznawano tylko *prawdy objawione*.

Przemiany społeczno polityczne epoki Odrodzenia doprowadziły do zerwania ciągłości z mrocznymi wiekami Średniowiecza. Powrócono wówczas do ideałów starożytnych, w tym także do racjonalnego poznawania rzeczywistości. Prace astronomiczne Galileusza stały się początkiem nowego przyrodoznawstwa. On jako pierwszy zastosował matematykę do opisu zjawisk fizycznych. Pokazał tym samym, że możliwa jest synteza poznania empirycznego z poznaniem rozumowym. Od tego momentu datuje się rozwój „obiektywnych” metod opisu przyrody. Według Galileusza przyroda jest pewnym uniwersalnym projektem napisanym w języku matematyki, a matematyka przekazuje pełną prawdę o rzeczywistości.

Dalszemu rozwojowi epistemologii towarzyszyła pogłębiająca się symbioza matematyki i nauk przyrodniczych. Znajduje ona wyraz w burzliwym rozwoju fizyki, będącej podstawą innych nauk przyrodniczych. W epoce Oświecenia wiara w nieograniczone możliwości racjonalnego poznania świata była tak duża, że wygłaszano przekonania o możliwości pełnej kontroli człowieka nad prawami przyrody. Wtedy też powrócono do sporu między poznaniem empirycznym a poznaniem racjonalnym. Dopiero rozważania Immanuela Kanta pozwoliły pogodzić oba stanowiska. Kant wykazał, że rejestrowane wyniki doświadczeń nie mają żadnej wartości, jeśli nie zostaną przetworzone przez umysł i zinterpretowane.

Po fascynacji w rozumne poznanie rzeczywistości, wiek XVIII przyniósł kolejne zwątpienie w możliwość zrozumienia przyrody. Rozbudzone nadzieje kompletnego poznania świata nie okazały się płonne. Jan Jakub Rousseau twierdził, że „natura chroni (ludzi) od wiedzy, tak jak matka wrywa niebezpieczną broń z ręki dziecka”. Dostrzeżono wówczas ograni-

czoność poznania naukowego. Dlatego też przełom wieku XVIII i XIX wiązał się z powrotem do zmysłowego i mistycznego poznawania natury, mimo dalszych odkryć w naukach przyrodniczych.

Dopiero wiek XIX i rewolucja techniczna odbudowała społeczną wiarę w sens nauki. Od tego wieku rozpoczyna się proces doskonalenia metod badawczych zapewniających obiektywizację opisu rzeczywistości.

Podstawą poznania naukowego na przełomie wieki XIX i XX były fakty empiryczne, z których metodą indukcji budowało się ogólne prawa. Ten sposób podejścia do poznania naukowego został zakwestionowany przez Karla Poppera. Wprowadził on tzw. *zasadę falsyfikowalności*. Poznanie naukowe jest zbiorem pewnych hipotez o otaczającym świecie, przy czym dana hipoteza jest naukowa tylko wtedy, gdy może być odrzucona przez wykonanie stosownego eksperymentu. Przykładowo, hipoteza o istnieniu Boga nie jest hipotezą naukową, gdyż nie można zaplanować żadnego eksperymentu, który w jednoznaczny sposób odrzucałby istnienie Boga.

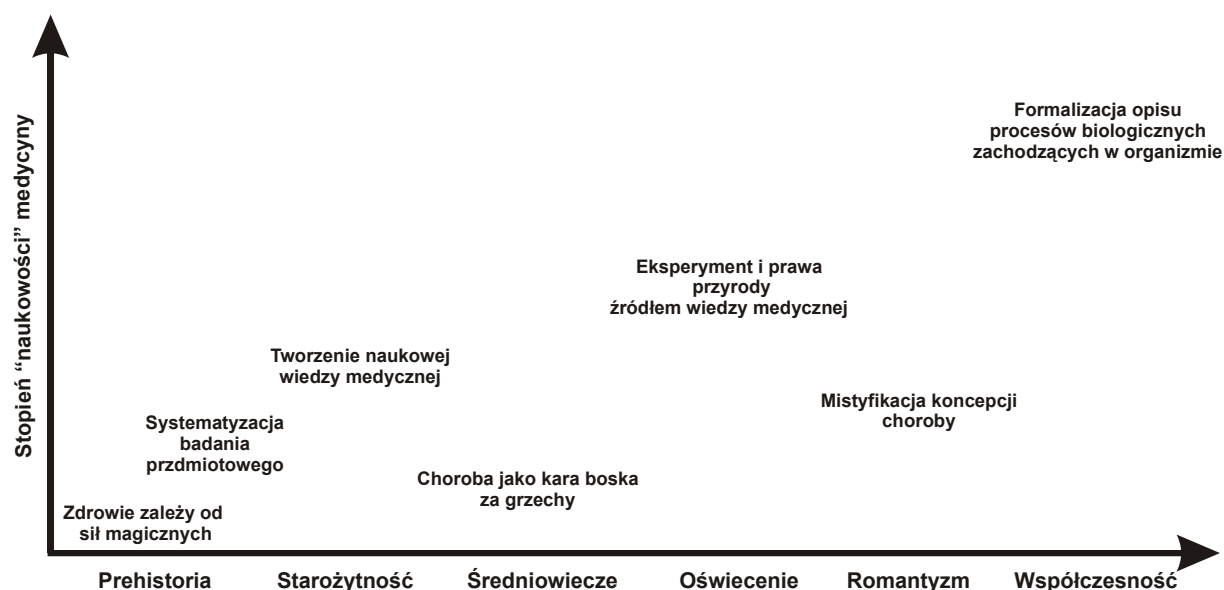
Koncepcja Poppera jest aktualna do dziś. Pomija jednak aspekty psychologiczne i społeczne poznania naukowego, które w dobie globalizacji świata odgrywają coraz większą rolę. Dlatego też warto wspomnieć o koncepcji Thomasa Khuna, który wprowadza pojęcie *paradygmatu*. Jest to zbiór pewnych twierdzeń, które zostają uznane za prawdziwe przez całe środowisko naukowców zajmujących się daną dziedziną. Koncepcja paradygmatów wprowadza pojęcie rewolucji poznania rzeczywistości. Jeżeli zostanie wykazane, że dany paradygmat jest sprzeczny z nowymi faktami, dokonuje się rewolucja w danej dziedzinie i nowy paradygmat zaczyna być obowiązującym. Wpływ środowiska naukowego na istniejące paradygmaty jest charakterystyczny dla nauk stosowanych, w tym także w medycynie. Pewną odmianą paradygmatu są uzgodnienia ekspertów na międzynarodowych konferencjach dotyczących standardów diagnostycznych i leczniczych.

Wprowadzenie do epistemologii paradygmatów uczyniło poznanie naukowe nieskończonym i niedoskonałym procesem dochodzenia do prawdy. Koncepcja ta jest zbliżona do pojęcia *modelu* obiektu, który ma być przedmiotem poznania. Pojęcie modelu obiektu zostanie wprowadzone w dalszej części skryptu.

## **2.2 Poznanie naukowe w medycynie**

Zmieniające się na przestrzeni epok koncepcje teorii poznania determinowały również sposób naukowego poznania zjawisk biologicznych, które warunkują zdrowie bądź chorobę.

Na rys. 2.1 przedstawiono charakterystyczne dla danej epoki koncepcje tworzenia wiedzy medycznej.



Rys. 2.1 Różnice w koncepcjach poznania naukowego w medycynie na przestrzeni epok

Anatomiczne i funkcjonalne uwarunkowania zdrowia oraz choroby budziły zawsze ciekawość człowieka. W czasach prehistorycznych, 2,5 tys. lat p.n.e. warunkiem zachowania zdrowia był sojusz z duchami, mocami tajemnymi. Jedyną metodą leczenia było uprawianie rytualnej magii mającej na celu wyproszenie łaski u duchów. Pomimo przekonania o władczej mocy bóstw, zachowały się zapiski na egipskich papirusach, które opisywały schemat prawzoru badania klinicznego. Już ok. 1200 lat p.n.e. zwracano uwagę na konieczność uważnego oglądania, osłuchiwania i badania palpacyjnego chorego. Obserwacje zebrane podczas badania były częścią modlitwy kierowanej do bóstw, bądź określały złe duchy, które należało wypędzić z organizmu. Jedną z metod diagnostycznych stosowanych w tym okresie było „czytanie choroby” z wyglądu wątroby zwierzęcia zabijanego w ofierze badanej osoby.

Pierwsze próby naukowego poznania procesów odpowiedzialnych za utrzymanie zdrowia rozpoczęły się w Starożytności. Żyjący na przełomie V i IV w. p.n.e. Hipokrates wprowadził pierwszą systematyzację klinicznego badania lekarskiego oraz wykazał konieczność prowadzenia historii choroby. Według pism Hipokratesa, podstawą do rozpoznania choroby jest obserwacja. Pojęcie zdrowia Hipokrates ograniczał do utrzymania stanu równo-



wagi między czterema sokami: krwią, śluzem, żółcią oraz czarną żółcią. Ważnym osiągnięciem Hipokratesa było opanowanie umiejętności prognozy przebiegu choroby.

Rozwój hipokratejskiej koncepcji uprawiania medycyny zaowocował w II w. n.e. opracowaniem pierwszej koncepcji, wyjaśniającej patogenезę chorób. Wprowadził ją grecki lekarz Galen, który przyczynę chorób upatrywał w zaburzeniach równowagi między czterema wymienionymi sokami. Zalecana przez Galena diagnostyka lekarska sprowadzała się do obserwacji ciepłoty i wilgotności ciała.

W II w. n.e. zaczęły powstawać pierwsze uniwersytety medyczne, które uczyły medycyny scholastycznej, czyli opartej na autorytecie wykładającego.

Okres Średniowiecza przyniósł znaczny regres w naukowym poznaniu medycyny. Na uniwersytetach, prowadzonych przez duchownych rozwinęła się tzw. *jatroteologiczna* koncepcja choroby. Zakładała ona, że zdrowie i choroba są całkowicie zależne od woli Boga. Wolę tę należy przyjmować z pełną pokorą, a jedyną formą terapii może być modlitwa. W wiekach średnich następowało mieszanie obserwacji przyrodniczych z ich interpretacją teologiczną. Począwszy od XII w. sobory i papieże wprowadzili dla duchownych zakaz zajmowania się medycyną. Jednak dalszym ciągu utrzymywano papieski protektorat nad europejskimi uniwersytetami, ograniczający powrót do starożytnej koncepcji poznania naukowego. Alternatywą dla medycyny jatroteologicznej było czarownictwo i magia, które zaczęły się dynamicznie rozwijać w drugiej połowie epoki średniowiecznej. Mimo, iż w socjologii religii praktyki religijne i magia traktowane są jako oddzielne formy wierzeń, efektywność leczenia prowadzonego przez ówczesnych duchownych i czarownice była podobna.

Mistyczne podejście do medycznych koncepcji epistemologicznych skończyło się wraz z zakończeniem Średniowiecza. Jednak w mentalności wielu osób średniowieczny sposób myślenia o zdrowiu i chorobie przetrwał aż do czasów współczesnych. Żywa jest wciąż w niektórych kręgach polskiej medycyny scholastyczna koncepcja poznania. Można przypuszczać, że wynika to z głęboko zakorzenionej w narodzie polskim filozofii chrześcijańskiej oraz tradycji romantycznych.

W epoce Odrodzenia nastąpił powrót do starożytnych metod poznawczych w medycynie. Od tej epoki rozpoczyna się powstanie medycyny naukowej opartej na eksperymencie i rozumie. Dokonujący się postęp w alchemii, chemii oraz fizyce zapoczątkował rozwój jatroteologii oraz jatrofizyki. W 1628 r. Harvey zaproponował pierwszy matematyczny opis układu krążenia, zaś w 1679 r. Borelli przedstawił biomechaniczny opis mięśni. Prekursorem mechanicznego opisu pozostałych procesów fizjologicznych był także Kartezjusz.

W epoce Renesansu rozwinął się także załazek współczesnej biometrii. W 1614 roku Santorio opublikował pracę *De statica medicina*, w której po raz pierwszy opisał zastosowanie pomiaru wagi, temperatury i ciśnienia tętniczego krwi.

Epoka Odrodzenia zapoczątkowała rozwój metodologii badań naukowych. Od tego momentu podstawą diagnostyki medycznej stają się przesłanki wynikające z wiedzy anatomicznej, fizjologicznej i patomorfologicznej. Upowszechnienie się badań autopsyjnych, a tym samym rozwój anatomii patologicznej prowadził do statycznego opisu choroby, zwanego opisem ontologicznym. Do XIX w. utożsamiano chorobę tylko ze zmianą patomorfologiczną, ignorując etiologiczną rolę zaburzeń czynnościowych. Dopiero rozwój eksperymentów fizjologicznych prowadzonych przez niemieckich naukowców zapoczątkował patofizjologiczne spojrzenie na chorobę jako proces zachodzący w czasie.

Tempo rozwoju naukowego poznania istoty zdrowia i choroby zostało zmniejszone w epoce Romantyzmu. Nastąpiło wówczas odejście od mechanistycznego opisu funkcjonowania organizmu w kierunku medycyny witalistycznej. Według tej koncepcji poznanie istoty chorób może być tylko metafizyczne, a źródłem choroby jest niepokój duszy. Pojawia się wówczas powrót do średniowiecznego, transcendentalnego ujęcia medycyny oraz roli Boga w procesie poznawczym i praktyce klinicznej. Jednak nie wszyscy ówczesni lekarze podzielali koncepcję medycznego mistycyzmu. W tym czasie również doskonalono metody badania chorego, czego przykładem jest wprowadzenie w 1819 r. przez Laenneca stetoskopu. Medycyna romantyczna, mimo że wiązała się z odejściem od zasad „obiektywnego” poznania ludzkiego organizmu, zwróciła uwagę badaczy na związek między stanem psychicznym a zaburzeniami somatycznymi organizmu. Poczynione wówczas rozważania epistemologiczne stały się podstawą do rozwoju przyszłej psychiatrii i medycyny psychosomatycznej.

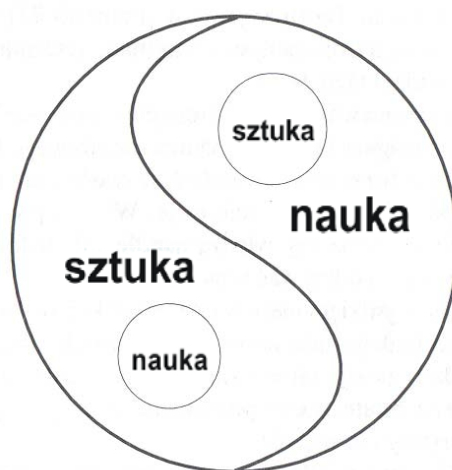
Od przełomu wieku XIX i XX do chwili obecnej obserwuje się wzrost stopnia sformalizowania metodologii badawczej stosowanej w medycynie. Realizacja tego trendu jest możliwa dzięki rozwojowi techniki (aparatura diagnostyczna, laboratoryjna) oraz matematycznych podstaw opisu procesów biologicznych (biomatematyka, biometria, biostatystyka). Ostatnim etapem ewolucji poznania naukowego w medycynie jest wprowadzona w 1980r. *medycyna oparta na dowodach*. Według tej koncepcji, podstawą poznania naukowego w medycynie jest prawidłowo metodologicznie zrealizowany eksperyment (zaprojektowany na podstawie teorii eksperymentu), właściwie dobrana metoda matematycznego opisu wyniku eksperymentu oraz umiejętna interpretacja biologiczna rezultatów analizy matematycznej. Opublikowane podobne wyniki z licznych eksperymentów tworzą podstawy naukowe do wy-

znaczenia standardów postępowania klinicznego, rekomendowanych przez środowiskowe towarzystwa naukowe.

Z uwagi na dominującą pozycję medycyny opartej na dowodach w obecnej praktyce klinicznej, zostanie ona szczegółowo omówiona w rozdziale 11. Tej koncepcji medycyny będzie także podporządkowana metodologia badań naukowych przedstawiana w skrypcie.

Dokonując porównania ewolucji teorii epistemologicznych i sposobu naukowego poznania w medycynie można zauważyć koherencję tych procesów historycznych. Ogólne podejście do rozumienia przyrody wywierało bardzo silny wpływ na epistemologię medyczną. Można także wykazać, że medycyna oparta na dowodach jest kompilacją zasady falsyfikacji Poppera oraz teorii paradygmatów Kuhna.

Przez wszystkie wieki stałym elementami w medycznej teorii poznania był empiryzm i racjonalizm. Obecnie medycyna wspiera się na dwóch filarach *nauce i sztuce*. Dualizm ten można przedstawić wykorzystując parafrazę konfucjańskiej symboliki *ying-yang* (rys.2,2). Poznanie naukowe jest wspierane sztuką tworzenia metodologii poznania. Z kolei sztuka podejmowania decyzji klinicznych korzysta z wiedzy naukowej.



Rys. 2.2 Symboliczne ujęcie współczesnej medycyny

### 3. OBSERWACJA JAKO ŹRÓDŁO DANYCH POMIAROWYCH

#### 3.1 Obserwacja jako źródło informacji

Percepcja informacji pochodzącej z otoczenia jest jedną z psychofizjologicznych czynności człowieka. Wykorzystuje ona wszystkie zmysły człowieka. Najwięcej informacji ze świata zewnętrznego, około 80% jest odbierana przez zmysł wzroku, a w następnej kolejności zmysł słuchu, dotyku, smaku. Zdecydowana większość informacji podlega kontroli autonomicznego układu nerwowego, zapewniając tym samym optymalne warunki do funkcjonowania człowieka w środowisku. Percepcja informacji może także być czynnością świadomie wymuszoną. Sytuację taką w języku potocznym nazywa się obserwacją. Przykładem takich sytuacji może być oglądanie filmu lub obserwowanie sytuacji na drodze podczas kierowania samochodem. Pomimo wolicjonalnego uruchomienia procesu obserwacji, w tych sytuacjach nie dokonujemy świadomej selekcji postrzeganych informacji. Dlatego w ten sposób rozumiana *obserwacja* nie jest *obserwacją naukową* ani *obserwacją kliniczną*.

Obserwację można uznać za *obserwację naukową* wtedy, gdy są ściśle określone następujące elementy składowe:

- Zbiór obiektów będącym przedmiotem obserwacji, oznaczany jako  $O = \{o_1 \ o_2 \ \dots \ o_n\}$ . Przez pojęcie obiekt rozumie się dowolny, ściśle ustalony fragment rzeczywistości, będącej przedmiotem badania. Obiektami mogą więc być pacjenci, zwierzęta laboratoryjne, wybrane procesy fizjologiczne lub patologiczne, wybrane elementy morfologiczne.
- Zbiór cech obiektów, oznaczany przez  $C = \{c_1 \ c_2 \ \dots \ c_m\}$ . Cecha określa atrybut opisujący dany obiekt, który jest istotny dla realizacji badania naukowego
- Metodę, bądź metody określające sposób uzyskania informacji o danej cesze. Najczęściej jest konieczność stosowania różnych metod dla różnych cech. W szczególności możemy potrzebować określić zbiór metod obserwacji oznaczany jako  $M = \{m_1 \ m_2 \ \dots \ m_k\}$ , gdzie  $k \leq m$
- Zbiór  $W = \{w_1 \ w_2 \ \dots \ w_m\}$  warunków, przy których musi być wykonana obserwacja. Przykładowo, warunkiem niezbędnym do prawidłowej obserwacji tętna pacjenta jest stan odprężenia psychofizycznego. Podobnie, warunkiem rzetelnej obserwacji postawy pacjenta, jest zapewnienie odpowiedniego oświetlenia w miejscu dokonywania obserwacji. Warunki obserwacji często wynikają z przyjętych metod ob-

serwacji. Mogą także wynikać z uwarunkowań badania naukowego, np. z ograniczeń bioetycznych.

- Obserwator – osoba lub osoby dokonujące obserwacji, oznaczane przez zbiór

$$R = \{r_1 \quad r_2 \quad \cdots \quad r_j\}. \text{ Zazwyczaj } j \text{ jest dużo mniejsze od } m.$$

Obserwacja naukowa<sup>1</sup> dotyczy więc konkretnych cech uprzednio wybranych obiektów przy zastosowaniu ustalonych metod pomiarowych w zadanych warunkach. Zdefiniowanie wszystkich elementów obserwacji tworzy *plan obserwacji*, który można oznaczyć jako  $P_o = \{O, C, M, W, R\}$ .

Jeżeli zbiór  $O$  składa się z pacjentów, zbiór  $C$  składa się z cech klinicznych opisujących stan pacjenta, dana obserwacja staje się *obserwacją kliniczną*.

Przy tak zdefiniowanym problemie obserwacji należy zwrócić uwagę na pojęcie *obserwowalności cechy obiektu*. Jeżeli dla danej cechy, przy zadanych warunkach obserwacji, nie można znaleźć metody pozyskania informacji o tej cenie, staje się ona *nieobserwowalna*. Należy też rozróżnić *nieobserwowalność bezwzględną* (niezależną od narzuconych warunków) od *nieobserwowalności względnej*, wynikającej z przyjętych ograniczeń. Dobrym przykładem ilustrującym nieobserwowalność względną może być ognisko endometriozy zlokalizowane na otrzewnej. Guzki endometriozy są nieobserwowalne w badaniu klinicznym i USG, natomiast można je zaobserwować podczas laparoskopii.

Dana cecha może być także nieobserwowalna względem metody obserwacji. Załóżmy, że należy dokonać obserwacji położenia topograficznego, wielkości oraz konsystencji narządów jamy brzusznej. Można to wykonać stosując dwie metody obserwacji: badanie palpacyjne oraz badanie ultrasonograficzne. Dwie pierwsze cechy są obserwowalne (najczęściej) w badaniu palpacyjnym i ultrasonograficznym. Precyzyjna informacja o konsystencji narządu może być dostępna tylko w badaniu USG.

Bardzo ważną właściwością każdej obserwacji jest jej subiektywizm. Wynik obserwacji ściśle zależy od wszystkich składników procesu obserwowania, w tym także od samego obserwatora. Na jakość obserwacji mają wpływ cechy osobowościowe obserwatora, jego kondycja psychofizyczna. Dlatego często zachodzi potrzeba wprowadzenia kilku obserwatorów dokonujących tej samej obserwacji. Otrzymanie zgodnych obserwacji dla co najmniej dwóch niezależnych obserwatorów zwiększa obiektywizm i wiarygodność obserwacji.

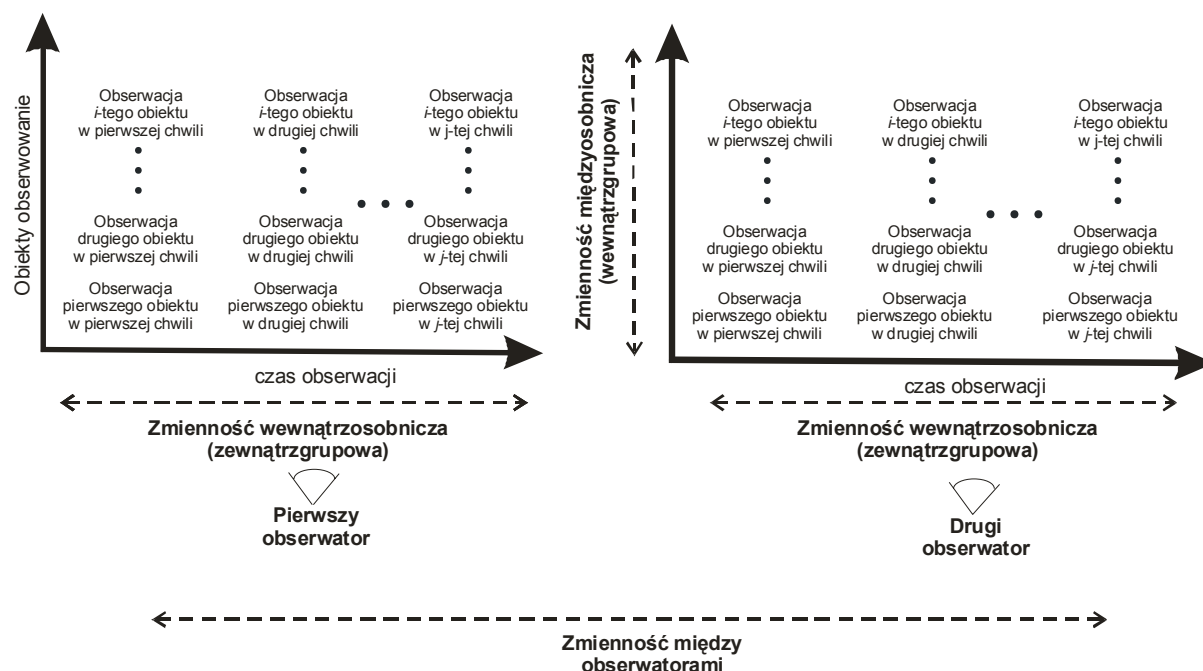
---

<sup>1</sup> Dla uproszczenia zapisu, w dalszej części skryptu każda obserwacja będzie traktowana jako obserwacja naukowa (lub kliniczna), stąd przymiotnik „naukowa” będzie opuszczany.

**Źródła zmienności obserwacji.** Każdy z elementów badania jest źródłem zmienności wyników obserwacji. Jednak dla przeprowadzenia rzetelnych obserwacji postuluje się, by podczas obserwacji zapewnić niezmienność metod i warunków obserwacji. Oznacza to, że dla wybranej cechy stosujemy dokładnie tę samą metodę i warunki do każdego obiektu obserwacji. Spełnienie postulatu niezmienności metody i warunków obserwacji zapewnia kolejną ważną właściwość – *powtarzalność obserwacji*. Powtórzenie obserwacji dla tych samych warunków oraz przy użyciu tej samej metody powinno dawać takie same wyniki obserwacji.

Źródła zmienności obserwacji, które są obecne w każdym procesie obserwacji, wynikają z różnic pomiędzy obiektami oraz różnic między chwilkami obserwacji (rys. 3.1). Różnica między obiektami wynika najczęściej z różnic osobniczych. Zarówno pacjenci, jak i zwierzęta (obiekty te występują najczęściej w medycynie) różnią się pod względem każdej cechy. Różnice te mogą wynikać z działania różnorodnych czynników, jak np. uwarunkowań genetycznych, warunków życia, przebytych zdarzeń wpływających na stan psychosomatyczny organizmu. Zmienność ta nazywana jest *zmiennością międzyosobniczą* lub *wewnątrzgrupową*.

Drugi typ zmienności obserwacji wynika z wpływu czasu na obserwowany obiekt. Często zmienność ta nazywana jest *zmiennością wewnątrzosobniczą*, gdyż dotyczy tego samego obiektu obserwowanego w różnych chwilach czasowych. Pamiętając o sentencji wypowiedzianej przez Heraklita, iż „wszystko płynie”, należy uwzględnić w planie obserwacji zmienność wewnątrzosobniczą. Oczywiście, można wskazać pewne obserwacje nie podlegające wpływowi czasu. Przykładem może być obserwacja wieku pierwszej miesiączki, wieku menopauzy u kobiet. Zmienność wewnątrzosobnicza także zależy od horyzontu czasowego obserwacji, dlatego często jest zanedbywana przy ich dokonywaniu. Jeżeli czas obserwacji obiektów jest krótszy od dynamiki zmian w nim zachodzących, można przyjąć, że w obserwacji nie występuje zmienność wewnątrzosobnicza.



**Rys. 3.1 Źródło zmienności obserwacji w medycynie**

W większości realizowanych badań naukowych także jest pomijana zmienność wynikająca z „wkładu” obserwatora w proces obserwacji. Można usprawiedliwić takie postępowanie wtedy, gdy wynikiem obserwacji jest wartość pomiaru dokonywanego przez urządzenie pomiarowe, a rola obserwatora ogranicza się do odczytania tej wartości. Uzasadnieniem tej tezy może być proste doświadczenie. Zadanie polega na obserwacji upływu czasu między dwoma zdarzeniami. Dla dwóch osób subiektywne poczucie czasu będzie inne. Co więcej, wyniki ostatnich badań wykazują, że subiektywne poczucie tempa upływającego czasu zmienia się z wiekiem obserwatora. Zastosowanie w tym prostym eksperymencie czasomierza sprawi, że wynik obserwacji długości danego odcinka czasu, dokonany przez dwie osoby będzie taki sam (z dokładnością odczytu pomiaru).

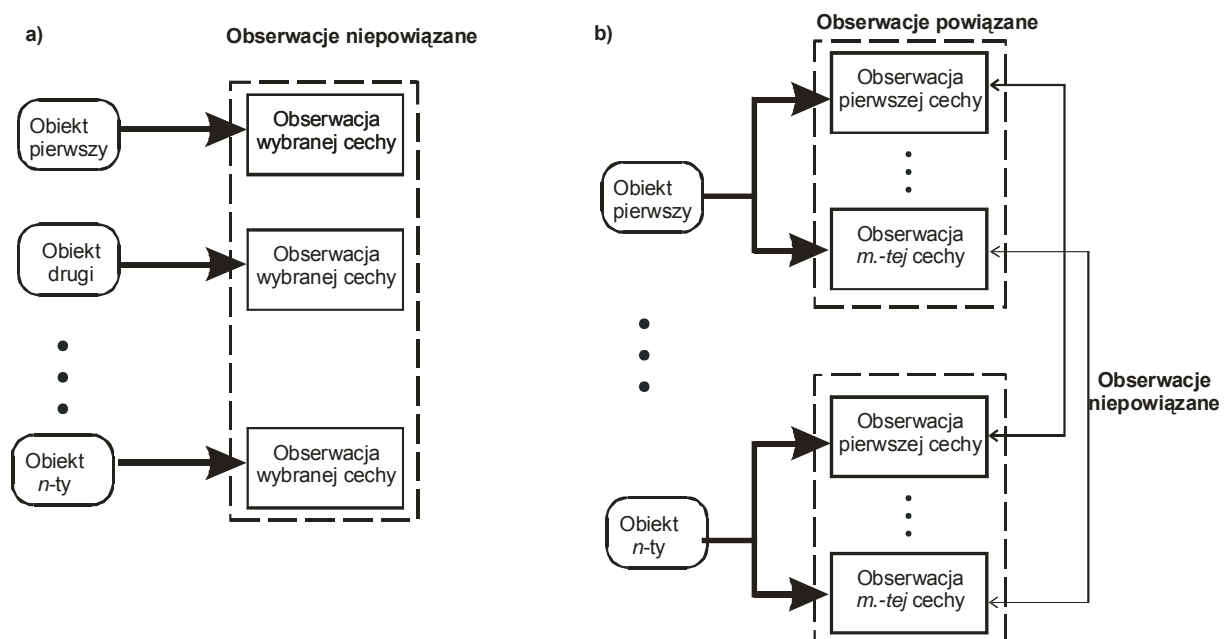
**Układ obserwacji.** Przez *układ obserwacji* należy rozumieć wzajemne powiązanie między obserwacjami oraz obiektami. Wyróżnia się układ obserwacji powiązanych i niepowiązanych. Obserwacje są *powiązane* wtedy, gdy można wskazać dowolny związek między obserwacjami. W przeciwnym wypadku, obserwacje są *niepowiązane*. Warto zauważyć, że znajomość charakteru powiązań między obserwacjami nie jest potrzebna do oceny ich powiązania.

Najprostszym układem obserwacji jest obserwowanie jednej cechy dla ustalonego obiektu. W badaniach naukowych, a także badaniu klinicznym sytuacja taka nie występuje. Pojedyncza obserwacja niesie bowiem zbyt mało informacji o badanym obiekcie. Dlatego w

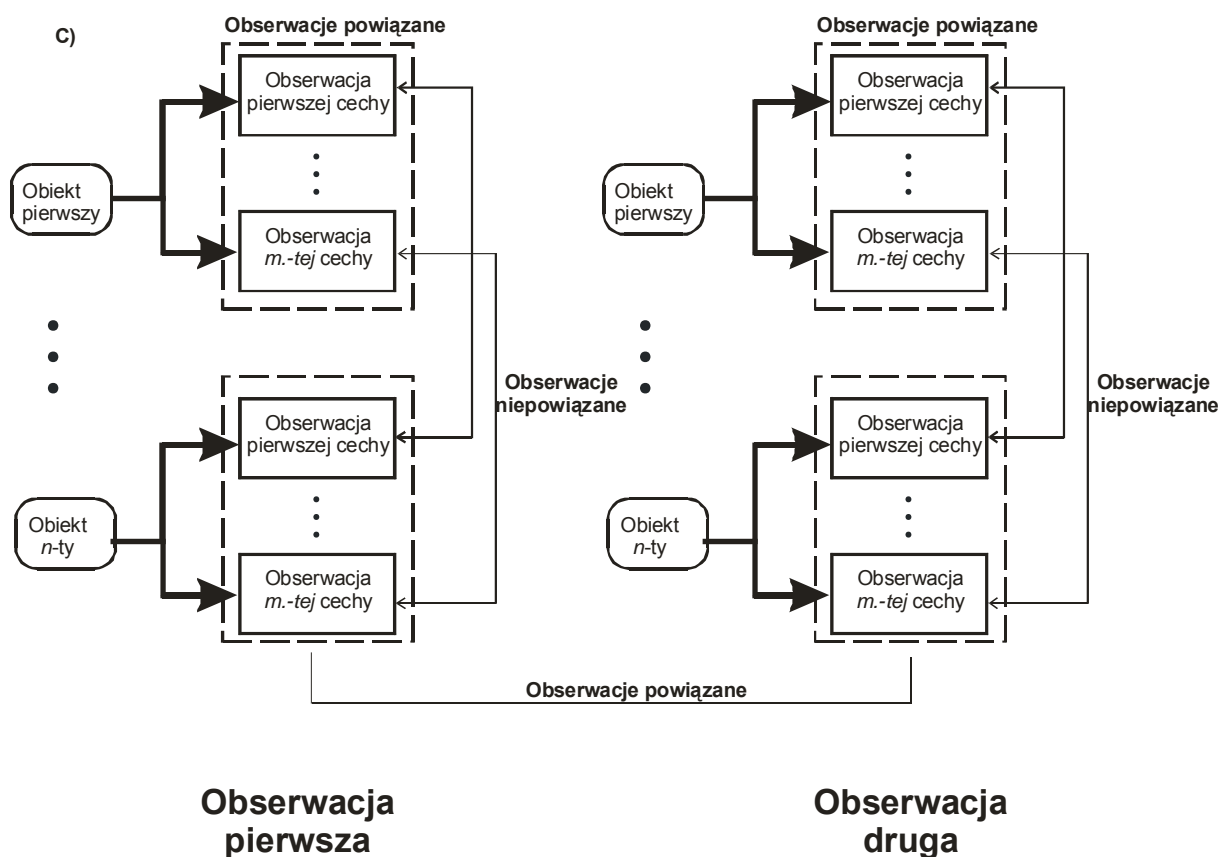
badaniach naukowych najczęściej mamy do czynienia z grupą pacjentów bądź zwierząt laboratoryjnych. W bardzo prostych eksperymentach obserwuje się wybraną cechę u każdego pacjenta. Jeżeli pacjenci zostali losowo przydzieleni do badanej grupy, a w szczególności nie są ze sobą spokrewnieni, otrzymujemy  $n$  niepowiązanych obserwacji. W tej sytuacji często używa się określenia *obserwacji jednowymiarowej*, gdyż dotyczy tylko jednej cechy.

Bardziej skomplikowany układ obserwacji otrzymuje się, gdy obserwuje się  $m$  cech u  $n$  pacjentów. Wtedy  $m$  obserwacji tworzy układ obserwacji powiązanych, gdyż dotyczą one tego samego pacjenta. Taką obserwację często nazywa się obserwacją  $m$ -wymiarową. Można zauważyć, że jeśli wybierzemy z zestawu  $m$  obserwacji wyniki otrzymane dla ustalonej cechy, wtedy otrzymuje się układ obserwacji niepowiązanych.

Kolejną komplikacją układu obserwacji jest jednoczesna obserwacja  $m$  cech u  $n$  pacjentów w różnych chwilach czasowych, np. przed podaniem leku i po podaniu leku. Ponieważ obserwacje w czasie dotyczą tych samych pacjentów, mamy układ obserwacji *podwójnie powiązanych*. Wszystkie trzy sytuacje ilustruje rys. 3.2.







Rys. 3.2 Układy obserwacji występujące w badaniach naukowych w medycynie ; a) układ obserwacji niepowiązanych b) układ obserwacji powiązanych c) układ obserwacji (podwójnie) powiązanych

### 3.2 Pomiar jako metoda obiektywizacji obserwacji

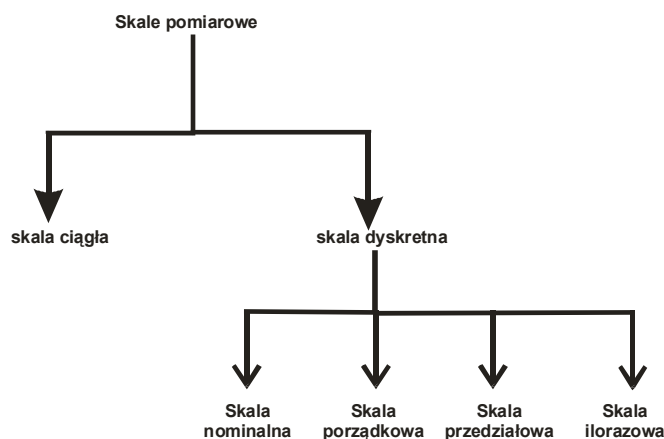
Subiektywizm wyników obserwacji bardzo ogranicza bezpośrednie ich wykorzystanie do obiektywnego poznania badanego obiektu. Uniemożliwia także dokonanie obiektywnych porównań między obserwacjami. Przykładowo, badając palpacyjnie gruczoł tarczycowy przed i po leczeniu możemy zauważyć zmianę jego objętości. Nie możemy jednak precyzyjnie wypowiedzieć się, jak duża jest to zmiana. Ponadto, małe zmiany objętości tarczycy mogą być zupełnie niezauważalne przez obserwatora. Dlatego już w XVII w Galileusz sformułował następującą zasadę; *Licz, co można policzyć, mierz co można zmierzyć, a to co nie jest mierzalne, uczyn mierzalnym*. Zasada ta również zaczęła obowiązywać w medycynie od czasu rozwoju metodologii badawczej nauk przyrodniczych. Z kolei rozwój medycyny opartej na dowodach zwiększył wykorzystanie pomiaru w praktyce klinicznej.

*Pomiarem* nazywa się takie przyporządkowanie symboli do obserwacji, które umożliwia porównywanie wyników obserwacji. Innymi słowy, pomiar wprowadza pewną relację porządku w zbiorze wyników obserwacji. Podczas pomiaru najczęściej przyporządkowuje się

im liczby. Taki pomiar jest najbardziej obiektywny, gdyż wykorzystuje dobrze zdefiniowaną na zbiorze liczb relację porządku. Pomiar może także wykorzystywać porządek leksykalny liter lub porządek semantyczny kategorii pojęciowych. Ten ostatni sposób pomiaru jest stosowany często w medycynie do cech pozornie niemierzalnych. Przykładem może być pomiar bólu, przyporządkowując natężeniu bólu kategorii {brak bólu, ból łagodny, ból ostry}. W metrologii pomiar, którego wynikiem są wartości liczbowe nazywa się  *pomiarem twardym(ściśłym)*. Odpowiednio  *pomiarem słabym* określa się pomiar polegający na przyporządkowaniu kategorii pojęciowych. Zbiór wszystkich możliwych wartości pomiaru nazywa się  *skalą pomiarową*.

Formalnie pomiar można zdefiniować następująco. Niech zbiór  $\Omega = \{\omega_1, \dots, \omega_n\}$  oznacza wyniki obserwacji dla ustalonej cechy, zaś zbiór  $S = \{s_1, s_2, \dots\}$  będzie skalą pomiaru. Pomiarom cechy  $\Omega$  nazywa się funkcję  $f$  przekształcającą zbiór  $\Omega$  w zbiór  $S$  o następującej właściwości: fakt:  $\omega_1 \neq \omega_2$  implikuje zależność  $s_1 \neq s_2$ . Najczęściej jednak nakłada się na funkcję  $f$  silniejszy warunek, aby z faktu  $\omega_1 \preceq \omega_2$  wynikał fakt  $s_1 \preceq s_2$ <sup>2</sup>.

**Skale pomiarowe.** W celu uporządkowania metodyki pomiaru i analizy wyników pomiarowych otrzymywanych w medycynie, wprowadzono następujący podział skal pomiarowych, przedstawiony na rys. 3.3



**Rys. 3.3 Podział skal pomiarowych**

Skala pomiarowa jest  *skalą ciągłą*, jeśli zbiór wartości pomiarów  $S$  może być reprezentowany przez zbiór liczb rzeczywistych. Skalę tę stosuje się do pomiaru cech uznawanych potocznie za „mieralne”, np. waga, wzrost, ciśnienie krwi, stężenie hormonu we krwi, itp.

<sup>2</sup> Znak „ $\preceq$ ” oznacza, że uporządkowanie zbioru może być wg dowolnego kryterium

Jeżeli zbiór  $S$  jest reprezentowany przez liczby całkowite lub kategorie pojęciowe, otrzymujemy *skale dyskretną*. Skalę tę stosuje się dla cech pozornie „niemierzalnych”. Można wyróżnić 4 typy skali dyskretnej.

Założmy, że zbiór  $S$  składa się z dwóch kategorii pojęciowych  $s_1$  i  $s_2$ , takich że można je odróżnić między sobą, ale nie można wskazać relacji porządku tzn.  $s_1 \overset{?}{\prec} s_2$ , wówczas otrzymujemy *skale nominalną*. Przykładem cech mierzonych w skali nominalnej jest płeć, wyznanie. Trudno bowiem wykazać, że płeć kobieca jest ważniejsza od płci męskiej, podobnie jak nie można dowieść, iż osoba wierząca jest „lepszą” od agnostyka.

Jeżeli jednak możliwe jest uporządkowanie pomiarów dokonanych w skali nominalnych, otrzymamy *skale porządkową*. Przykładem skali porządkowej jest pomiar natężenia bólu  $S = \{\text{brak bólu, ból łagodny, ból ostry}\}$ .

Jeżeli dla wyników pomiarowych możemy obliczyć różnice między wynikami, która ma sens medyczny, otrzymujemy wówczas pomiary w *skali przedziałowej*. Charakterystyczną właściwością tej skali jest możliwość umownego przyjęcia początku skali (punktu zerowego). Zdarza się to wtedy, gdy istnieją różne metody pomiaru tej samej obserwacji, np. pomiar stężenia hormonu w IU/ml oraz w mg/l. Mimo wzajemnego przesunięcia w wartościach tych skal, w obu przypadkach różnice między oznaczonymi stężeniami mają to samo znaczenie biologiczne.

Ostatnim typem skali jest *skala ilorazowa*, która ma te same właściwości co skala przedziałowa oraz umożliwia zachowanie stosunku między pomiarami. Skala ilorazowa najczęściej jest reprezentowana przez pomiar w skali ciągłej, aczkolwiek może być stosowana dla odpowiednio dobranych skal dyskretnych. Każdy pomiar „mierzalnych” cech odbywa się w skali ilorazowej, np. pomiar wieku, czasu, wagi, stężenia substancji w osoczu krwi.

W zależności od zastosowanej skali pomiar reprezentuje różną ilość informacji o mierzonej cesze. Najmniej informacji dostarcza skala nominalna, najwięcej skala ilorazowa. Skalę dokładniejszą można zawsze zastąpić skalą mniej dokładną, choć w ten sposób traci się część informacji (rys. 3.4). Dlatego przy projektowaniu procedury pomiarowej zaleca się wybierać skalę dającą najwięcej informacji o obserwowanej cesze.



Rys. 3.4 Wzajemna relacja między skalami pomiarowymi

### 3.3 Pomiary wykorzystywane w medycynie

Pomiary wykorzystywane w medycynie obejmują następujące kategorie obserwacji: pomiar cech biofizycznych obiektu, pomiar biologicznych i biochemicznych składników płynów ustrojowych oraz psychologicznych i socjologicznych determinantów zdrowia i choroby

**Pomiar cech biofizycznych.** Z uwagi na możliwość pomiaru cech biofizycznych obiektu w skali interwałowej, wykonuje się je przy wykorzystaniu urządzeń pomiarowych. Typowymi wielkościami biofizycznymi mierzonymi w medycynie są:

- cechy morfometryczne - długość, wzrost, waga, objętość, cechy kształtu, np. pomiar łuku wysklepienia stopy, pomiar kąta określającego stopień skrzywienia bocznego kręgosłupa
- cechy kinetyczne i dynamiczne –zmiany położenia kończyn, szybkość ruchu kończyn, siła napięcia mięśni, wielkość drżenia (tremoru), siła skurczu mięśnia sercowego, macicy, prędkość przepływu, ciśnienie
- właściwości bioelektryczne i biomagnetyczne tkanek albo narządów – oporność skóry, natężenie pola elektrycznego wytwarzanego przez serce, mózg, pole magnetyczne wytwarzane przez serce, mózg.

W zależności od mierzonej cechy i obserwowanego obiektu, do pomiaru cech biofizycznych stosuje się proste narzędzia (np. waga, sfigmomanometr) bądź urządzenia zaawansowane technologicznie (np. aparat USG, jądrowy rezonans magnetyczny).

**Pomiar biologicznych i biochemicznych składników płynów ustrojowych.** Pomiary tych cech mogą wykorzystywać wszystkie skale pomiarowe w zależności od danej cechy. Doko-

nywane są za pomocą nowoczesnej aparatury, która wykorzystuje właściwości biologiczne, biochemiczne lub biofizyczne mierzonej cechy. Przykładem mogą być urządzenia laboratoryjne służące do oznaczeń wybranych substancji w osoczu krwi bądź spektrometry masowe.

**Pomiar psychologicznych i socjologicznych determinantów zdrowia i choroby.** Pomimo, że cechy te uznawane są potocznie za „niemierzalne”, pomiar psychologicznych i socjologicznych uwarunkowań zdrowia oraz choroby odgrywa coraz większą rolę we współczesnej medycynie. Obserwacja tych cech jest szczególnie ważna przy opisie epidemiologicznego modelu danej choroby, co będzie przedstawione w następnych rozdziałach. Głównym narzędziem stosowanym do pomiaru tych cech są testy psychologiczne, bądź specjalne konstruowane kwestionariusze. W odróżnieniu od poprzednich pomiarów, dla których metody pomiarowe były „ukryte” w aparaturze, w tej grupie można wyróżnić dwie metody pomiarowe. Pierwsza z nich polega na wykorzystaniu obserwatora – ankietera do przeprowadzenia wywiadu, druga polega na „autobserwacji” - pacjent sam wypełnia ankietę. Każda z tych metod ma wady i zalety, które zostaną zasygnalizowane w następnym rozdziale. Psychologiczne i socjologiczne determinanty zdrowia i choroby najczęściej są mierzone w skali nominalnej lub porządkowej.

### **3.4 Obserwacja i pomiar w praktyce klinicznej**

Początkowo teoria obserwacji i pomiaru była stosowana w medycynie tylko w eksperymentach naukowych. Jednak opisane zasady prowadzenia obserwacji i pomiaru stosują się również do obserwacji klinicznej pacjenta. Ostatnio powstała nowa gałąź wiedzy medycznej zwana *biometrią kliniczną* (ang. *clinometrics*).

Zauważmy bowiem, że badanie podmiotowe i przedmiotowe pacjenta odpowiada dokonywaniu bezpośrednich bądź pośrednich obserwacji i pomiarów wybranych cech klinicznych. Wszystkie elementy badania klinicznego, czyli wywiad, osłuchanie, obmacywanie i opukiwanie są *implicite* pewnym pomiarem. Osłuchując bowiem serce pacjenta, obserwowaną cechą są tony serca, zaś wynikiem „pomiaru” jest częstotliwość tonów oraz ich miarowość. Z kolei rodzajem pomiaru przy osłuchiwaniu płuc jest subiektywna wartość częstotliwości i natężenia szmerów oddechowych. Podobnie, badanie palpacyjne narządów wewnętrznych pozwala dokonać pomiaru ich położenia topograficznego, wielkości, spoistości, a niekiedy także lokalizacji i natężenia bólu.

Pomiary te dokonywane są w różnych skalach pomiarowych: od skali nominalnej stosowanej przy wywiadzie do skali ilorazowej, stosowanej przy badaniach dodatkowych.

Taki sposób spojrzenia na badanie kliniczne prowadzi do następujących wniosków praktycznych. Po pierwsze, skoro badanie podmiotowe i przedmiotowe jest rodzajem obserwacji, można stosować dla niego plan obserwacji, wymuszający określenie elementów obserwacji. Zatem, podczas wywiadu dobrze jest najpierw określić cel zadawania pytania, tak by odpowiedź pacjenta stanowiła obserwacje (informacje) dotyczącą wybranej cechy. Należy także dążyć, do tego by pomiary cech klinicznych były dokonywane na możliwie najbardziej dokładniej skali pomiarowej. Ilustruje to następujący przykład. Badając palpacyjnie jamę brzuszną można zadać pacjentowi pytanie „*Czy Pana tu boli?*”. Pytanie to jest metodą pomiaru bólu w skali nominalnej. Wprowadzając jednak skalę porządkową bólu pytaniem „*Czy tu boli bardziej?*”, możemy dokładniej określić przestrzenny rozkład natężenia bólu. Niektóre objawy wymagają wręcz stosowania skali porządkowej. Sytuacją taką jest obserwacja objawu Blumberga, gdy natężenie bólu jest większe przy oderwaniu dłoni badającego od powłok brzusznych niż przy ucisku. Pomiar objawu Blumberga przy wykorzystaniu skali nominalnej nie dostarczy lekarzowi żadnej informacji o stanie pacjenta.

Po drugie, o ile warunki przeprowadzania badania klinicznego pozwalają, należy dbać, by obserwacja i pomiar odbywały się zawsze w tych samych warunkach, przy zastosowaniu tych samych metod obserwacji. Zalecenie to prowadzi do standaryzacji przeprowadzania obserwacji klinicznej, niezbędnej przy porównywaniu wyników obserwacji.

Po trzecie, dokonując obserwacji klinicznej, pacjenta należy zawsze pamiętać o źródłach zmienności obserwacji. W szczególności ważne jest uwzględnianie zmienności pochodzącej od obserwatora, czyli osoby wykonującej badanie. Wyniki badania klinicznego, podobnie jak wyniki każdego pomiaru zawsze są obarczone pewnym *błędem pomiarowym* oraz *niepewnością* otrzymanego wyniku. Istota błędu pomiarowego zostanie omówiona w następnym rozdziale, Fakt niepewności wyników badania pacjenta powinien inspirować do zachowania pokory wobec uzyskanych wyników, a także zachęcać do przeprowadzenia badania przez inną osobę w przypadku braku przekonania o ich „jednoznaczności”.

Obserwacje i pomiary kliniczne mają także cechy odróżniające ich specyfikę od obserwacji naukowej. Ograniczenia ekonomiczne, a także czasowe często nie pozwalają na wykonanie wszystkich możliwych pomiarów w danej sytuacji klinicznej. Z drugiej zaś strony, rozwój możliwości diagnostyki klinicznej prowadzi do wielowymiarowej struktury obserwacji klinicznej, co może utrudniać syntezę informacji zawartej w wynikach badania.

Podsumowując, klinicysta musi być zawsze świadomym, że obserwacja kliniczna pacjenta może być *niepewna*, *niekompletna* oraz *wielowymiarowa*. Atrybuty te powinny być uwzględniane w procesie stawiania diagnoz lekarskich.

## 4. BŁĄD POMIAROWY I NIEPEWNOŚĆ WYNIKÓW W BADANIACH NAUKOWYCH I PRAKTYCE KLINICZNEJ

### 4.1 Definicja błędu pomiarowego

W poprzednim rozdziale określiliśmy elementy składowe procesu obserwacji i pomiaru. Zarówno subiektywizm obserwatora, zastosowane warunki oraz metody obserwacji i pomiaru powodują, że wynik pomiaru zawsze odbiega od wartości rzeczywistej. Wartość ta jest najczęściej niemożliwa do zmierzenia. Dlatego wynik pomiaru należy traktować tylko jako przybliżenie „prawdziwej” wartości, dokonane z pewnym błędem.

W celu wyjaśnienia pojęcia błędu pomiarowego, załóżmy, że znamy wartość rzeczywistą pomiaru, oznaczoną symbolem  $x$ . Przez  $\hat{x}$  oznaczmy wartość otrzymaną w wyniku pomiaru. Różnica między wartością zmierzoną a wartością rzeczywistą nazywana jest *błędem pomiarowym wielkości  $X$* , który będzie oznaczany przez  $\delta X$ . Błąd pomiarowy wyraża się wzorem:

$$\delta X = \hat{x} - x \quad (4.1)$$

Ze wzoru tego wynika, że im błąd pomiarowy jest bliższy zeru, tym dokładniejszy jest wynik pomiaru. Z kolei znak błędu pomiaru informuje obserwatora o przeszacowaniu (gdy  $\delta X > 0$ ) lub niedoszacowaniu (gdy  $\delta X < 0$ ) otrzymanych wyników.

Mimo, że wartość rzeczywista mierzonej cechy najczęściej jest nieznana, błąd pomiarowy odgrywa ważną rolę w biometrii. Po pierwsze, pozwala on porównywać pomiary otrzymane różnymi metodami. Wówczas wartość pomiaru otrzymana dokładniejszym narzędziem pomiarowym jest traktowana jako wartość rzeczywista. Często występują sytuacje, w których im jest dokładniejsza procedura pomiarowa, tym większe są koszty wykonywania pomiarów. Dlatego można wówczas skorzystać z narzędzia pomiarowego o mniejszej dokładności, będąc świadom wartości popełnianego błędu.

Drugim zastosowaniem błędu pomiarowego jest problem *kalibracji* urządzenia lub procedury pomiarowej. W wielu sytuacjach dokładność urządzenia pomiarowego jest zależna od regulowanych parametrów. Wówczas zakłada się wartość dopuszczalnego błędu pomiarowego i tak dobiera się parametry, by pomiar był obciążony co najwyżej przyjętym błędem. Przykładem takiej sytuacji może być pomiar stężenia leku w osoczu na chromatografie. Dokładność tego pomiaru zależy od stężeń i składu chemicznego rozpuszczalnika.

Trzecim typowym zastosowaniem błędu pomiarowego jest badanie powtarzalności danej procedury pomiarowej. Najczęściej wymaga się, by przy zachowaniu jednakowego pla-

nu obserwacji dokonywanej w różnych momentach czasowych, błąd pomiaru był podobny. Oczywiście jest to możliwe przy pomiarze cech niezmiennych (lub wolnozmiennych) w czasie.

Błąd pomiarowy, określony wzorem (4.1) nazywany jest *błędem bezwzględnym*. Ten sposób oceny dokładności nie odnosi się do samej wartości mierzonej i jest powszechnie stosowany tam, gdzie zakres mierzonej wielkości jest niewielki. Wiele cech posiada jednak dużą rozpiętość wartości mierzonych i wówczas trudno jest utrzymać tę samą wartość błędu bezwzględnego. Załóżmy dla przykładu, że dysponujemy metodą pomiarową o błędzie bezwzględnym równym 1. Jeśli zmierzymy wartość 20 to błąd pomiarowy będzie stanowił 5% wartości zmierzonej i często można taką sytuację zaakceptować. Jeśli jednak zmierzymy wartość równą 1 to błąd będzie stanowił dużą część wartości zmierzonej, w tym przypadku będzie to 100% i na ogół pomiar taki jest bezużyteczny.

Dlatego do oceny dokładności pomiaru używa się *błędu względnego*, który wyraża się następującym wzorem:

$$\delta_w X = \frac{\hat{x} - x}{x} 100\% = \frac{\delta X}{x} 100\%, \text{ dla } x \neq 0 \quad (4.2)$$

Błąd względny, liczony względem wartości rzeczywistej, staje się niezależny od niej. Stwierdzenie dla danej procedury pomiarowej błędu względnego równego 10%, oznacza, że każda wartość zmierzona będzie stanowiła 10% wartości rzeczywistej.

## 4.2 Źródła błędów pomiarowych

Źródłem błędu pomiaru może być: metoda pomiaru (obserwacji), warunki pomiaru (obserwacji) oraz obserwator. Przeanalizujemy to na przykładzie najczęściej dokonywanego w warunkach klinicznych pomiaru – ciśnienia tętniczego krwi metodą Korotkowa. Przykładem błędu wynikającego z metody może być wpływ ortostatycznej komponenty ciśnienia podczas szybkiej pionizacji pacjenta. Metoda Korotkowa nie uwzględnia w pomiarze ciśnienia tętniczego odpowiedzi z baroreceptorów, co może prowadzić do zafałszowania zmierzonej wartości ciśnienia.

Drugim źródłem błędów pomiaru mogą być warunki wykonywania pomiaru. Ich wpływ na wartość mierzonego ciśnienia jest znany pod nazwą „efektu białego fartucha”. Obserwuje się wówczas zawyżoną wartość mierzonego ciśnienia wynikającą z emocji towarzyszących badaniu lekarskiemu. Innym przykładem błędu wprowadzanego przez warunki pomiarowe może być wzrost ciśnienia wynikający z pokonania schodów przez pacjenta. Dlatego



też zawsze postuluje się, by ciśnienie tętnicze było mierzone po okresie habituacji układu krążenia pacjenta do warunków spoczynkowych.

Źródłem błędów pomiarowych może być także obserwator. W rozważanym przykładzie jedną z przyczyn tego błędu może być niedokładne odczytanie wskazania sfigmomanometru. Uważa się, że przyrządy rtęciowe pozwalają na dokładniejszy odczyt niż zegarowe.

Błąd obserwatora może też wynikać z różnej wrażliwości i charakterystyki częstotliwościowej słuchu osoby wykonującej badanie. Właściwości te decydują o chwili percepcji tonów Korotkowa, co przekłada się na zmierzone wartości ciśnienia skurczowego i rozkurczowego.

Wyniki ostatnich badań sugerują także, że płeć osoby badanej może prowadzić do zawyżonego pomiaru ciśnienia tętniczego. Fakt ten zaobserwowano u pacjentek wtedy, gdy badającymi lekarzami byli mężczyźni.

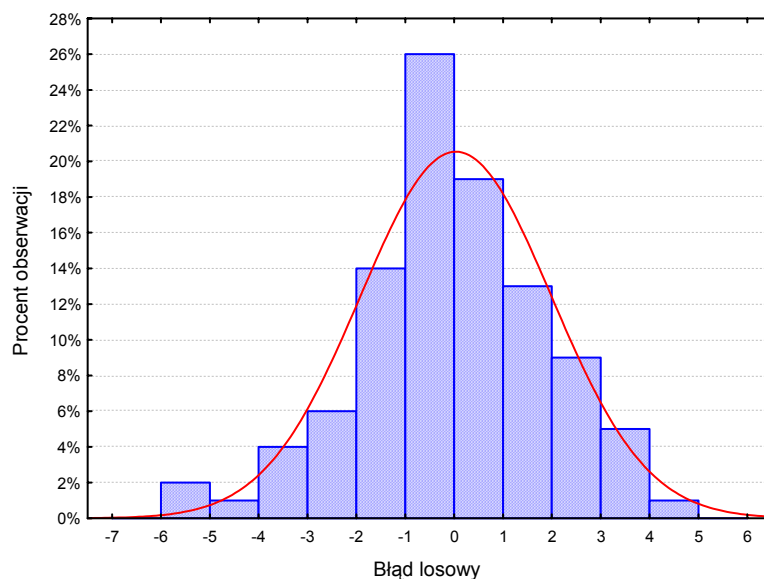
Pomiar ciśnienia tętniczego krwi w kontekście omawiania źródeł błędów pomiarowych został wybrany ze względów dydaktycznych. Odpowiednie przykłady czynników zakłócających wynik mierzenia można podać dla każdego pomiaru biometrycznego. Świadomość istnienia tych błędów jest niezbędna do prawidłowej interpretacji lub korekty otrzymywanych wyników.

#### **4.3 Klasyfikacja błędów pomiarowych**

Ze względu na charakter występowania błędu pomiarowego wyróżnia się: błędy systematyczne oraz błędy losowe (przypadkowe). *Błędem systematycznym* nazywamy błąd, który w sposób zdeterminowany powtarza się przy każdym pomiarze. Błędy systematyczne najczęściej mają tę samą wartość podczas powtarzania pomiaru tej samej cechy. Stały błąd systematyczny często jest nazywany *obciążeniem pomiaru*. Błędy systematyczne mogą pojawiać się w pewnym momencie procesu pomiarowego. Może to być wolna fluktuacja właściwości przyrządu pomiarowego i warunków wykonywania pomiaru lub zmiana skokowa wywołana na przykład wstrząsem mechanicznym.

Drugim typem błędów pomiarowych są błędy losowe. Błąd pomiaru jest *błędem losowym* wtedy, gdy jego wartość zmienia się w każdym pomiarze w sposób przypadkowy. W odróżnieniu od błędu systematycznego, zmiana ta dotyczy również znaku błędu. Źródłem błędów losowych są nieznane czynniki zakłócające pomiar, dlatego najczęściej nie są możliwe do całkowitego wyeliminowania. Ich obecność może wynikać np. z wpływu warunków atmosferycznych na stan psychosomatyczny pacjenta. Ponieważ nie są możliwe do określenia

dokładne wartości błędu losowego, opisuje się go zmienną losową. Błędy losowe najczęściej mają rozkład zbliżony do rozkładu normalnego (rys. 4.1).



**Rys. 4.1** Typowy rozkład częstości błędów losowych

Błędy losowe mają tę właściwość, że ich wartości w kolejnych pomiarach wahają się oscylując wokół zera. Średnia arytmetyczna z dużej liczby powtórzonych pomiarów daje w wyniku redukcję błędu losowego i jest to uznana metoda poprawiania dokładności. Wartość średnia w dużej liczbie powtórzonych pomiarów różni się od wartości rzeczywistej w zasadzie tylko o wartość błędu systematycznego.

Jeżeli przez  $\varepsilon_n$  oznaczmy wartość błędu dla  $n$ -tego pomiaru, to estymator wartości średniej i odchylenia standardowego tego błędu wyraża się wzorem

$$\begin{aligned}\bar{\varepsilon} &= \frac{1}{N} \sum_{n=1}^N \varepsilon_n \\ \sigma_{\varepsilon} &= \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\varepsilon_n - \bar{\varepsilon})^2}\end{aligned}\tag{4.3}$$

gdzie  $N$  jest liczbą wykonanych pomiarów. Wówczas  $\bar{\varepsilon}$  średnie jest oszacowaniem błędu systematycznego, a wartość standardowego odchylenia błędu określa rozrzut losowy błędu pomiarowego.

Jeżeli w pomiarze danej cechy występują błędy losowe, to najczęściej mają one następujące cechy: 1) są addytywne względem wartości „rzeczywistej” oraz 2) ich wartość średnia wynosi zero. Warunek addytywności oznacza, że wynik pomiaru jest sumą wartości rzeczywistej i błędu losowego, czyli  $\hat{x} = x + \varepsilon$ . Jedną z metod pozbycia się wpływu błędu losowego

o zerowej wartości jest uśrednienie wielu pomiarów. Otrzymujemy wówczas następującą zależność

$$\hat{\bar{x}} = \bar{x} + \bar{\varepsilon} \xrightarrow{\bar{\varepsilon}=0} \hat{\bar{x}} = \bar{x} \quad (4.4)$$

Powyższy podział błędów pomiarowych jest stosowany do wszystkich pomiarów, niezależnie od dziedziny naukowej. Pozwala on uzyskać odpowiedź na pytanie, czy występowanie danego błędu znacząco obciąża wyniki pomiaru oraz czy można się tego pozbyć.

W badaniach medycznych, głównie stosowanych w epidemiologii wyróżnia się jeszcze dwa inne rodzaje błędów pomiarowych: błędy zróżnicowane oraz niezróżnicowane (ang. *differential errors*). Błąd pomiarowy jest *błędem zróżnicowanym*, jeśli jego wartości lub rozkład zależy od wybranego czynnika. Błędy nie spełniające tego warunku nazywa się *błędami niezróżnicowanymi*. Wielkość błędów zróżnicowanych zależy od tego, czy badana osoba jest chora czy zdrowa. Dokładność oceny nasilenia objawów będzie inna u pacjentów chorych niż u osób zdrowych.

Inny przykład błędu zróżnicowanego związany jest z wykorzystaniem kwestionariusza do pomiaru wybranych cech psychometrycznych pacjenta. Jeżeli pytania w kwestionariuszu dotyczą odległej przeszłości, mamy do czynienia z *obciążeniem niepamięcią*. Załóżmy, że chcemy zbadać wpływ stabilizacji cykli miesięczkowych w okresie dojrzewania na występowanie zaburzeń płodności. Kobiety, które miały od menarche problemy z regularnością cyklu miesięczkowego będą, lepiej pamiętały przebieg swego dojrzewania niż kobiety, u których zaburzenia pojawiły się w wieku dorosłym. Wynikać to może także z faktu, że kobiety pierwszej grupy mogły częściej rozpocząć wizyty u ginekologa, a tym samym częściej były pytane o przebieg okresu dojrzewania. Ponadto u tych pacjentek można uzyskać mniej obciążone obserwacje z historii chorób dotyczących adolescencji.

Określenie czy dany błąd ma charakter zróżnicowany czy niezróżnicowany ma znaczenie podczas korekcji błędu pomiarowego.

#### 4.4 Zmienność obserwacji a błąd pomiarowy

Omawiając źródła błędów pomiarowych wskazaliśmy, że jednym z nich może być zmienność obserwacji. Jednak rozstrzygnięcie czy zmienność jest źródłem informacji czy źródłem błędu, zależy od celu wykonania pomiaru. Zilustrujemy to następującym przykładem.

W omawianym przykładzie dotyczącym pomiaru średniego stężenia estrogenów u kobiet, przy pobieraniu próbek krwi w różnych dniach cyklu miesięczkowego prowadzi do

systematycznego błędu losowego. Jeżeli jednak celem pracy byłoby określenie dynamiki zmian stężenia estrogenów w cyklu pacjentek nieplodnych, zmienność wewnątrzsobnicza obserwacji byłaby istotną informacją.

Podobnie dla określenia stężenia trójjodotyroniny w populacji pacjentów chorych na chorobę Gravesa Basedowa informacja o rozrzucie (dyspersji) tych stężeń zawiera użyteczne informacje. Z kolei gdybyśmy spróbowali na podstawie uzyskanych wyników populacyjnych przewidzieć stężenie hormonu u danego pacjenta, wartość dyspersji określałaby błąd wartości spodziewanej przy pomiarze.

Z powyższych przykładów wynika, że już cel przeprowadzanego badania (naukowego bądź klinicznego) ściśle determinuje sposób podejścia do błędów pomiarowych.

#### 4.5 Błędy pomiarowe a niepewność wyników

Błędy i obciążenia obserwacji i pomiarów sprawiają, że otrzymywane wyniki nigdy nie są pewne. Dlatego przy opisywaniu wyników pomiarów należy podawać wartość pomiaru oraz zakres jego niepewności. *Niepewność* pomiaru jest określana przez maksymalny błąd pomiaru. Jeżeli pomiar cechy  $X$  był obciążony błędami  $\delta_1 X, \dots, \delta_k X$ , to niepewność pomiaru określa wyrażenie  $U = \max\{|\delta_1 X|, \dots, |\delta_k X|\}$ . Najczęstszym sposobem prezentacji niepewności pomiaru jest określenie przedziału  $U$ , w którym znalazłoby się 95% potencjalnych wyników pomiarowych. Przedział taki nazywany jest *przedziałem ufności*.

Alternatywnym sposobem określania niepewności jest uzupełnienie wartości pomiaru o wartość miary niepewności, tzn, wynik pomiaru zapisujemy w postaci  $\hat{x} \pm u$ . Przykładem miary niepewności może być wariancja losowego błędu pomiarowego,  $\sigma_\varepsilon^2$ , lub standardowego odchylenia błędu będąca pierwiastkiem kwadratowym wariancji.

#### 4.6 Analiza rzetelności wyników obserwacji i pomiarów

Nadrzędnym celem, który powinien towarzyszyć każdej obserwacji i pomiarowi jest uzyskanie wiarygodnej informacji o mierzonej cesze. Dlatego powinny one przebiegać w ten sposób, aby maksymalizować rzetelność otrzymywanych wyników (co jest równoważne minimalizacji błędów pomiarowych).

Jeżeli pomiar jest wykonywany standardową metodą w optymalnych dla niej warunkach, oszacowanie rzetelności dokonuje się na podstawie danych udostępnionych przez twórców metody. Przykładowo, każde urządzenie pomiarowe ma określany graniczny błąd, z którym mierzy się obserwowane cechy. Podobnie każdy standaryzowany test psychometryczny lub socjometryczny ma określony przedział ufności dla osiągniętych wyników.

Problem jednak pojawia się wtedy, gdy do pomiaru wybranych cech należy opracować własną metodę pomiarową. Sytuacja taka rzadko dotyczy technicznych urządzeń pomiarowych, gdyż analiza i kontrola błędu pomiarowego jest stałym etapem projektowania i produkcji urządzenia. W trakcie jego eksploatacji niezbędna jest jedynie kontrola zachowania stałości wartości błędu lub okresowa kalibracja urządzenia.

Najczęściej problem ten dotyczy badań epidemiologicznych, w których należy użyć wyspecjalizowanego kwestionariusza. Drugą sytuacją mogą być badania kliniczne, w których kluczową rolę odgrywa obserwacja nowych objawów klinicznych. Wówczas przed wykorzystaniem wyników pomiarów do dalszego opracowywania, należy obowiązkowo wykonać analizę rzetelności. Z uwagi na obszerność metodologii analizy rzetelności, ograniczymy się tylko do przedstawienia idei oraz wybranych metod.

Analiza rzetelności sprowadza się do sprawdzenia relacji między wynikami otrzymywanymi nową metodą, a wynikami otrzymywanymi tzw. metodą referencyjną. Przez *metodę referencyjną* rozumie się taki sposób pomiaru danej cechy, dla którego znana jest dokładność pomiarowa. Jeżeli wyniki nowej metody, są porównywalne z wynikami dla metody referencyjnej, można uznać nową metodę za rzetelną. Potrzeba zastosowania nowej metody może wiązać się z jej uproszczeniem, obniżeniem kosztów lub dostosowaniem do możliwości wykonania w warunkach klinicznych.

Nie zawsze jednak jest dostępna metoda referencyjna. Wówczas o rzetelności pomiarów można wnioskować pośrednio. Jeżeli np. dwie nowe, różne metody pomiaru dają prawie takie same wyniki, możemy wnioskować, że mierzymy właściwą informację. Modyfikacja tej metody polega na pomiarze tej samej cechy w różnych momentach czasowych, przy założeniu niezmienności wewnątrzsobniczej. Zgodność otrzymanych wyników sugeruje dużą rzetelność metody pomiaru. Przykładowo, jeśli pytamy pacjentki o ich wiek menarche w trakcie dwóch wizyt oddalonych od siebie w czasie i otrzymujemy zbliżone wyniki, możemy wnioskować, że pacjentki podawały prawdziwy wiek pierwszej miesiączki.

Do oceny zgodności wyników pomiarowych można stosować dwa parametry. Dla wyników mierzonych w skali co najmniej przedziałowej stosuje się *współczynnik korelacji liniowej*. Załóżmy, że dysponujemy  $n$  wartościami pomiarów zarejestrowanych metodą refe-

rencyjną  $R_1, \dots, R_n$  oraz nową (testowaną) metodą pomiarową  $T_1, \dots, T_n$ . Wówczas współczynnik korelacji liniowej wyraża się wzorem:

$$r_{RT} = \frac{1}{n} \sum_{i=1}^n \left( \frac{R_i - \bar{R}}{\sigma_R} \right) \left( \frac{T_i - \bar{T}}{\sigma_T} \right) \quad (4.5)$$

gdzie  $\sigma_R, \sigma_T$  oznaczają odchylenie standardowe pomiarów dla metody referencyjnej i testowanej, zaś  $\bar{R}, \bar{T}$  są wartościami średnimi tych pomiarów. Im wartość  $r_{RT}$  jest bliższa jedności, tym obie metody dają wyniki pomiarowe bardziej zgodne.

Odpowiednikiem zgodności pomiarów mierzonych w skali nominalnej jest współczynnik kappa. Załóżmy, że mierzymy dwoma metodami cechę, którą można opisać dwoma kategoriami. Niech to będzie płeć osób określana za pomocą badania przedmiotowego i badania chromosomalnego (określającego kariotyp męski albo żeński). Grupując odpowiednio pacjentów uzyskamy cztery podgrupy pacjentów. Będą to kobiety i mężczyźni, dla których płeć chromosomalna jest tożsama z płcią gonadalną oraz 2 grupy kobiet i mężczyzn, dla których płeć gonadalna jest różna od płci chromosomalnej.

Uzyskane wyniki można przedstawić w tabeli przedstawiającej rozkład liczebności pacjentów w poszczególnych podgrupach (tab. 4.1).

**Tab. 4.1 Rozkład częstości pacjentów przy określaniu płci metodą badania somatycznego i chromosomalnego**

	Płeć żeńska	Płeć męska	Suma
Kariotyp XX	$p_{11}$	$p_{12}$	$P_{1\cdot}$
Kariotyp XY	$p_{21}$	$p_{22}$	$P_{2\cdot}$
Suma	$P_{\cdot 1}$	$P_{\cdot 2}$	1

Symbole  $p_{11}, p_{12}, p_{21}, p_{22}$  oznaczają odsetki pacjentów (prawdopodobieństwa) w poszczególnych podgrupach. Dla powyższej tabeli możemy zdefiniować *współczynnik zgodności kappa* wyrażający się wzorem

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (4.6)$$

gdzie  $P_o = p_{11} + p_{22}$ , zaś  $P_e = P_{\cdot 1}P_{1\cdot} + P_{\cdot 2}P_{2\cdot}$ . Im wartość  $\kappa$  jest bliższa jedności, tym bardziej dwie metody pomiaru dają zgodne wyniki.

Jak widać z powyższych rozważań, przeprowadzenie poprawnej analizy rzetelności wyników otrzymywaną nową metodą pomiarową wymaga zaprojektowania i przeprowadzenia dodatkowego badania. Dlatego, jeśli jest to możliwe powinno stosować się takie metody pomiarowe, których rzetelność została już oszacowana. Oszczędza to czas i koszty realizacji badania naukowego.

## 5. ZDROWIE I CHOROBA JAKO PRZEDMIOT POZNANIA NAUKOWEGO

Wykonanie pomiarów i obserwacji dotyczących badanego obiektu (zjawiska, pacjenta) jest tylko wstępnym etapem poznania naukowego w medycynie. Najważniejszym etapem poznania jest wyekstrahowanie wiedzy zawartej w obserwacjach poprzez znalezienie relacji między tymi obserwacjami. Uogólniając rozważania, można powiedzieć, że w medycynie mamy dwa podmioty poznania naukowego. Interesuje nas odkrycie prawidłowej morfologii i czynności organizmu ludzkiego, a także wyjaśnienie mechanizmów powstawania choroby. Zgodnie z obowiązującą definicją wprowadzoną przez WHO, *zdrowiem* określamy kompletny, fizyczny, psychiczny oraz społeczny dobrostan człowieka. Zatem celem poznania naukowego, w odniesieniu do zdrowia, będzie znalezienie czynników i mechanizmów biologicznych, psychicznych oraz społecznych, które pozwalają osiągnąć i utrzymać zdrowie.

Z kolei przez *chorobę* można rozumieć zaburzenie stanu dynamicznej równowagi organizmu człowieka. Wówczas przedmiotem poznania choroby jest jej etiologia oraz towarzyszące chorobie zmiany morfologiczne i funkcjonalne. Poznanie, a więc zrozumienie etiopatogenezy choroby jest niezbędnym warunkiem do opracowania skutecznych metod diagnostyki i leczenia. Dalsze rozważania w tym rozdziale będą koncentrowały się właśnie na poznaniu istoty choroby, choć bez trudu można je także stosować do poznania fenomenu zdrowia.

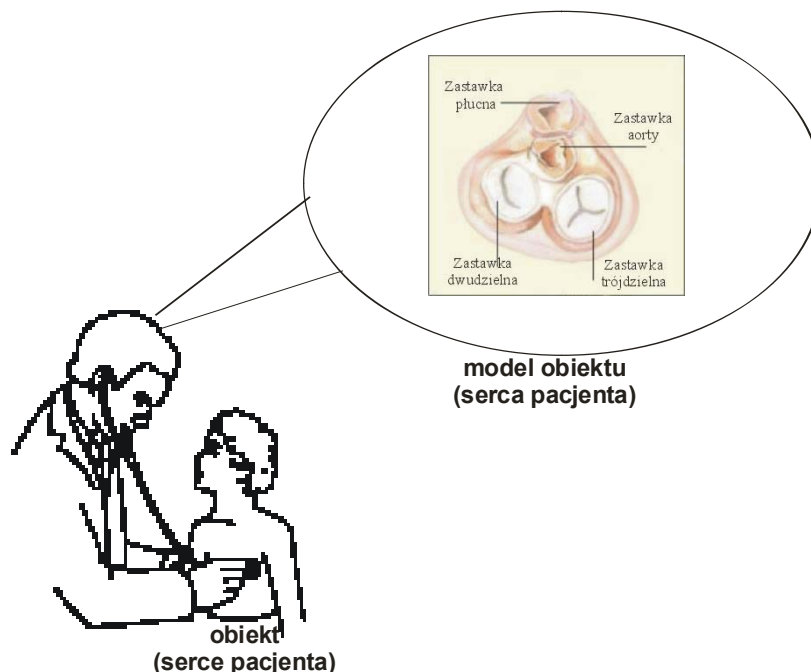
W pierwszym rozdziale skryptu przedstawiliśmy historyczne ujęcie poznania naukowego w medycynie. Konkluzją tych rozważań było wprowadzenie pojęcia *modelu obiektu*, będącego efektem procesu poznawczego. W rozdziale tym dokładnie omówimy to pojęcie i przedstawimy na czym polega modelowanie w naukach medycznych.

### 5.1 Pojęcie modelu procesu chorobowego

Rozważmy następującą sytuację. Załóżmy, że chcemy poznać obwodowy mechanizm prowadzący do rozwoju nadciśnienia tętniczego krwi. W tym celu dokonujemy pomiaru rozkurczowego ciśnienia tętniczego oraz pomiaru oporu naczynia oporowego u 20 wybranych losowo pacjentów. Sama rejestracja zestawu 20 par pomiarów nie wyjaśni nam wpływu zwiększonego oporu obwodowego na wartość ciśnienia tętniczego. Dopiero „wyłuskanie” z obserwacji, jaka jest relacja między wartością ciśnienia a wartością oporu pozwala znaleźć opis tłumaczący wybrany fragment patofizjologii nadciśnienia tętniczego. Otrzymujemy w ten sposób pewne wyobrażenie o badanym zjawisku, określane mianem *modelu*. Pojęcie *model* pochodzi od łacińskich słów *modus*, *modulus* i oznacza wzorzec, makietę. W rozumieniu



używanym w medycynie model stanowi formalny opis relacji między obserwacjami, reprezentując w przybliżony sposób, wiedzę o badanym obiekcie. Zatem model możemy traktować jako „uproszczoną” kopię wycinka badanej rzeczywistości. Schematycznie, relacje między modelem, obserwatorem i obiektem przedstawia rys. 5.2



Rys. 5.2 Relacja między obiektem, obserwatorem i modelem

Przykład ten pokazuje, że nawet podczas praktyki klinicznej, lekarz *implicite* posługuje się modelami. Przykładowo, osłuchiwanie tonów serca pacjenta pozwala stworzyć w umyśle lekarza model serca pacjenta. Model ten reprezentuje tylko wybrane cechy anatomiczne serca (np. prawidłowa budowa zastawek serca). Ponadto każdy model z definicji odwzorowuje rzeczywistość z pewną *niedokładnością*. Dlatego również stworzony w ten sposób model nie dostarcza „pełnej” wiedzy o budowie i czynności serca pacjenta i często musi być wzbogacony informacją z badań obrazowych.

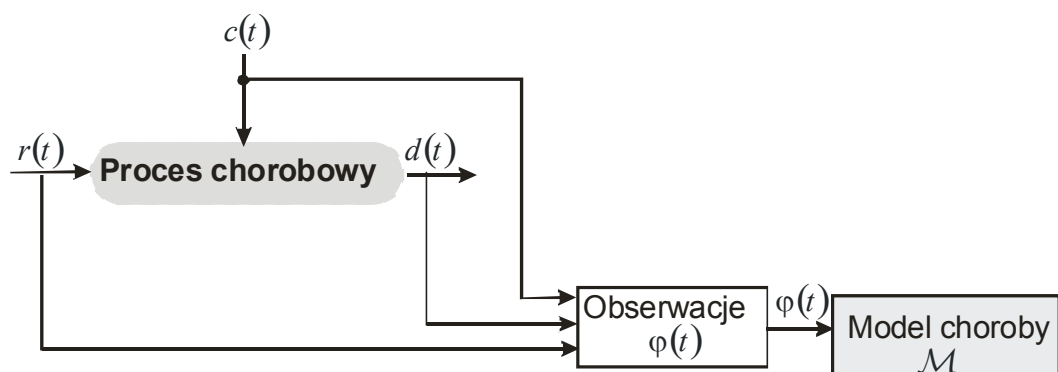
Modele procesów chorobowych, które spotykane są w medycynie występują w trzech postaciach. Najczęstszą formą są *modele werbalne*, które opisują słownie relację między badanymi czynnikami. Zaletą tych modeli jest duża dostępność dla środowiska medycznego, wadą zaś mała możliwości weryfikacji „prawdziwości” modelu. Drugą postacią są *modele graficzne*, w których relacje między obserwacjami reprezentowane są w formie wykresów lub tabel. Modele graficzne umożliwiają precyzyjniejszy opis relacji między obserwacjami niż modele werbalne.

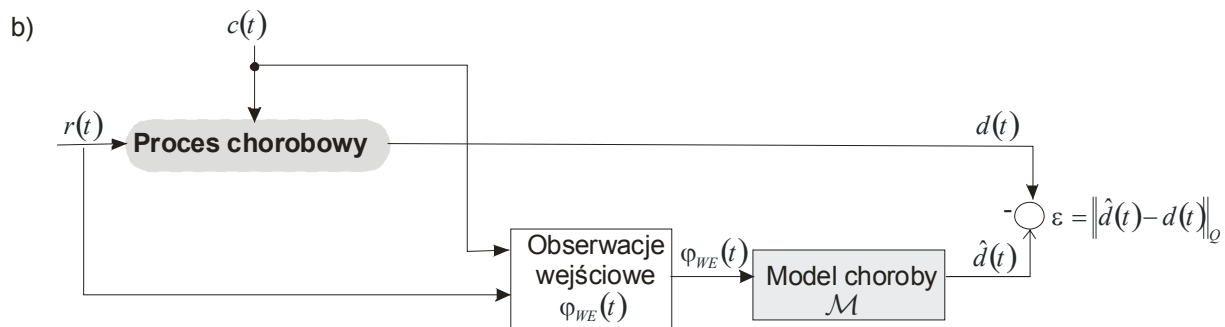
Najbardziej ścisły opis badanego procesu chorobowego dostarcza *model matematyczny*, w którym relacje między obserwacjami wyrażone są za pomocą równań matematycznych. Z uwagi na rosnące zastosowanie modeli matematycznych w medycynie, zostaną one omówione dokładniej.

### 5.1.1 Model matematyczny procesu chorobowego

Założmy, że celem przeprowadzanego eksperymentu jest zbadanie czy wybrany czynnik wpływa na rozwój badanej choroby. Dla przykładu, badaną chorobą będzie niepłodność wynikająca z niedojrzałości pęcherzyka jajnikowego, zaś badanym czynnikiem stężenie gonadotropiny FSH. Niech wykładnikiem (surogatem) choroby będzie średnica pęcherzyka Graafa, mierzona za pomocą USG. Ponieważ zarówno stężenie FSH, jak i średnica pęcherzyka jajnikowego zmieniają się w czasie założmy, że obserwacje będą zależne od czasu. Oznaczmy przez  $r(t)$  stężenie hormonu w chwili  $t$ , zaś przez  $d(t)$  zmierzoną średnicę pęcherzyka jajnika. Relacja między gonadotropiną a średnicą pęcherzyka może być zakłócana wpływem innych czynników, które należy też uwzględnić przy konstrukcji modelu danego procesu. W rozważanym przykładzie czynnikiem będzie masa ciała, która też może zmieniać się w czasie, a jednocześnie wiadomo, że wpływa na czynność hormonalną jajnika. Oznaczmy więc przez  $c(t)$  pomiar masy ciała pacjentki w chwili  $t$ . Otrzymujemy w ten sposób zestaw obserwacji, nazywany często wektorem obserwacji złożony z trzech elementów. Wektor taki często jest zapisywany w następujący sposób:  $\varphi(t) = [r(t) \ d(t) \ c(t)]$ . Jest on wykorzystywany do budowy modelu matematycznego (rys.5.3a).

a)





Rys. 5.3 a) Model procesu chorobowego b) Weryfikacja modelu

Zgodnie z rys. 5.3 modelem matematycznym badanego procesu chorobowego, oznaczany przez  $\mathcal{M}$  będziemy nazywać równanie

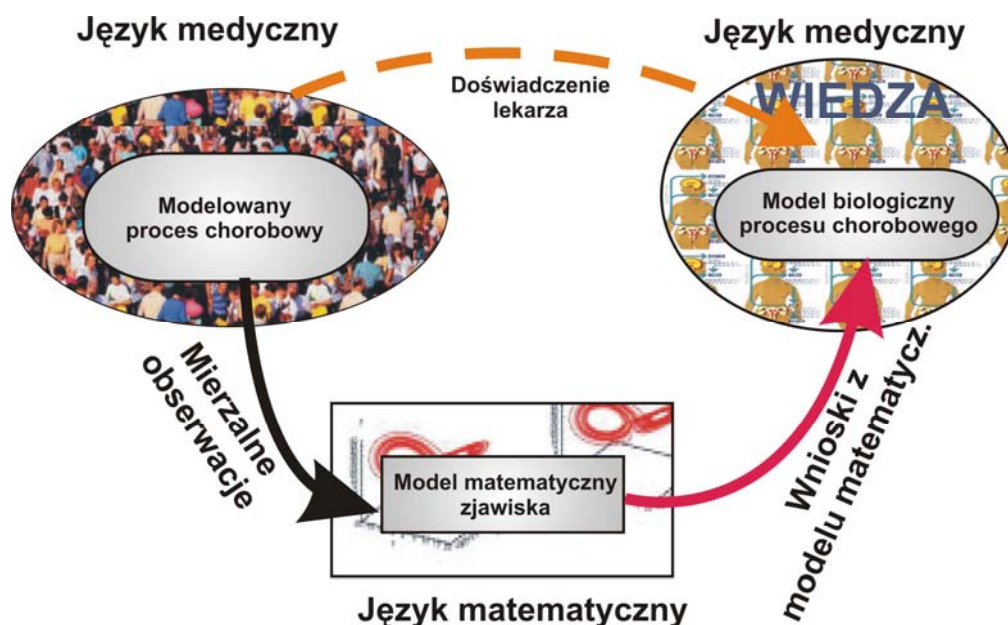
$$\hat{d}(t) = M(\varphi_{WE}(t); \theta), \quad (4.7)$$

które pozwoli „odtworzyć” modelowany proces na podstawie wektora obserwacji wejściowych  $\varphi_{WE}(t) = [r(t) \ c(t)]$ . W rozważanym przykładzie, model matematyczny powinien dać możliwość przewidzenia średnicy pęcherzyka Graafa  $\hat{d}(t)$  na podstawie obserwacji stężenia FSH oraz masy ciała pacjentki. Model też zawsze zależy od postaci opisującej relacje między obserwacjami, oznaczanej przez  $M$  i zwanej *strukturą modelu*. Drugim elementem każdego modelu jest zbiór jego parametrów, oznaczany przez  $\theta$ . Struktura modelu najczęściej jest dobierana *a priori*, czasem na podstawie wyników poprzednich badań. Parametry modelu są wyznaczane na podstawie obserwacji podczas *identyfikacji modelu*. Najczęstszym kryterium doboru parametrów jest żądanie, aby model możliwie najlepiej odtwarzał rzeczywistość, a więc postulujemy by wynik modelu  $\hat{d}(t)$  był „podobny” do rzeczywistej obserwacji  $d(t)$ . Im mniejsza jest różnica między modelem a rzeczywistością, tym otrzymujemy dokładniejszy model. Podobieństwo to można definiować na różne sposoby w zależności od przyjętego kryterium podobieństwa, oznaczanego przez  $Q$ . Różnica  $\varepsilon = \|\hat{d}(t) - d(t)\|_Q$  wyznaczana względem uprzednio przyjętego kryterium  $Q$  nazywana jest *błędem modelu*, który szacuje *niepewność modelu*. Źródłem niepewności modelu są:

- błędy pomiarowe obserwacji oraz ich niekompletność. Wybór składników wektora obserwacji jest zawsze arbitralny.
- ograniczona wiedza teoretyczna wykorzystywana do budowy modelu.
- przyjęte założenia, będące podstawą budowanego modelu.

Z przedstawionego rozumowania wynika bardzo ważny wniosek dotyczący modeli. Dla wybranego obiektu będącego celem modelowania, można znaleźć nieskończenie wiele modeli, a wybór modelu „najlepszego” jest czysto arbitralny. Możemy bowiem żądać, by model dobrze opisywał wybrany proces choroby u „średniego pacjenta”, albo u pacjenta najczęściej występującego w danej populacji lub u konkretnego, aktualnie badanego pacjenta. Ponadto modele mogą się różnić w zależności od przyjętego wektora obserwacji. Im więcej czynników uwzględniamy w modelu, tym (na ogół) otrzymujemy dokładniejszy opis modelowanego zjawiska. Odnosząc tę zasadę do naszego przykładu można spodziewać się, że uwzględnienie w modelu obserwacji stężenia pozostałych hormonów wpływających na dojrzewanie pęcherzyka Graafa, pozwoli skonstruować dokładniejszy model wyjaśniający mechanizm zaburzonego wzrostu pęcherzyka.

Można zatem powiedzieć, że tworzeniu modelu matematycznego wybranego procesu chorobowego towarzyszy transformacja wiedzy biomedycznej na język matematyczny, a po skonstruowaniu modelu, interpretacja biomedyczna wniosków wynikających z modelu. Dlatego też modelowanie matematyczne w medycynie może być traktowane jako metoda tworzenia wiedzy medycznej, uzupełniająca doświadczenie lekarza (Rys. 5.4)



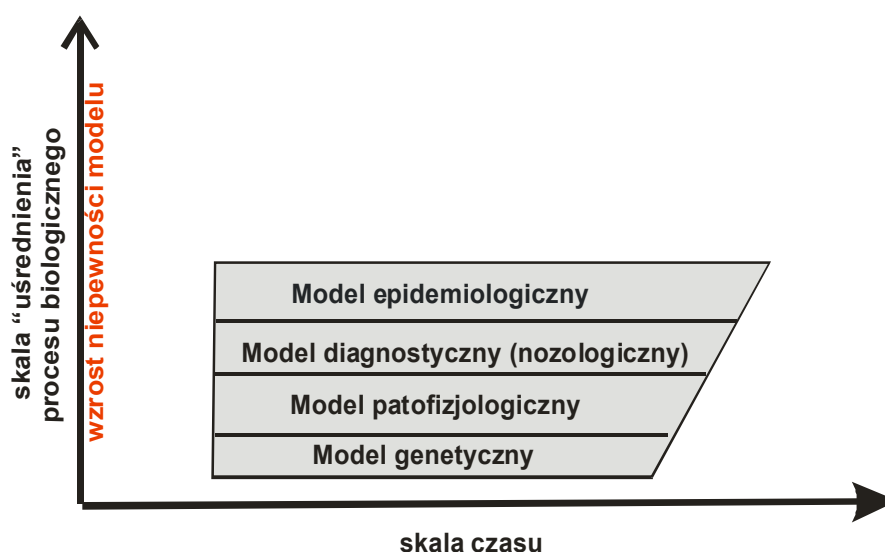
Rys. 5.4 Model matematyczny jako narzędzie tworzenia wiedzy medycznej

## 5.2 Hierarchiczny opis procesu chorobowego

Z uwagi na duży wysoki stopień złożoności procesów chorobowych, nie jest możliwe zbudowanie „całościowego” modelu opisującego patomechanizm danego procesu chorobowego. Dlatego też wydaje się zasadne wprowadzenie umownego parametru, określającego granulację modelu, czyli poziom szczegółowości modelu. Innymi słowy, granulacja modelu określa liczbę szczegółów „widzianych” przez osobę budującą model. Im większa granulacja, tym większe czasowo-przestrzenne uśrednienie modelowanego procesu.

Rozumowanie to pozwala rozpatrywać dany model w dwuwymiarowej przestrzeni granulacja  $\times$  czas, wprowadzając kolejne poziomy (warstwy) szczegółowości modelu (rys. 5.5).

Modelem najbardziej dokładnym opisującym patofizjologię choroby jest poziom molekularno genetyczny. Uśrednienie w czasie i przestrzeni zmian molekularnych prowadzi do zaburzeń czynnościowych na poziomie narządów, a zatem otrzymujemy model patofizjologiczny. Z kolei długotrwałe zaburzenia organiczno-czynnościowe powodują konkretne dolegliwości, dające się opisać modelem diagnostycznym danej choroby.



Rys. 5.5 Hierarchiczne modelowanie procesu chorobowego

Najbardziej uśrednionym modelem jest model epidemiologiczny. Na tym poziomie szukamy czynników ryzyka danej choroby, a więc proces chorobowy (najczęściej) jest modelowany przez dychotomiczną zmienną losową. Uśrednienie odbywa się także w czasie, gdyż np. stałe czasowe zmian czynnościowych są większe od stałych reakcji biochemicznych, etc.

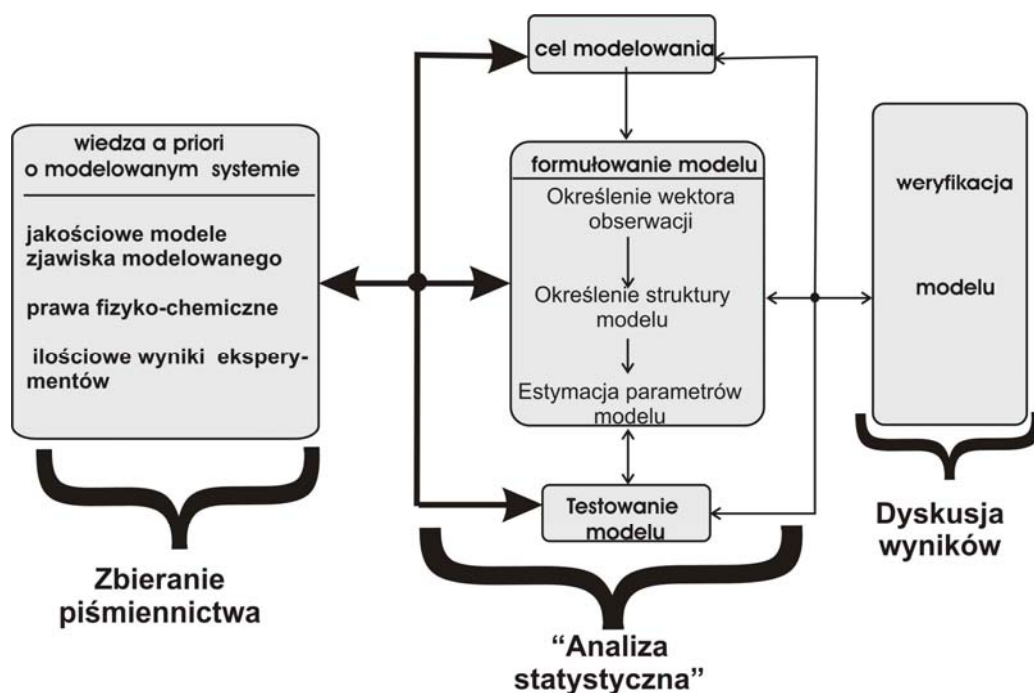
Zauważmy także, że im większa skala uśrednienia modelowanego procesu, tym większy wzrost niepewności tworzonego modelu. Widać to choćby po liczebności obserwacji niezbędnej na każdym poziomie modelowania, największa jest dla modeli epidemiologicznych.

Przedstawiona koncepcja podejścia do modelowania procesu chorobowego pozwala na analizę spójności różnych modeli, gdyż model na danym poziomie powinien wynikać (a co najmniej nie być sprzeczny) z modelem z poziomu niższego.

Z drugiej zaś strony, pomaga nam wybrać zestaw obserwacji niezbędny do opisu modelu na danym poziomie uszczegółowienia.

### 5.3 Tworzenie modelu badanego procesu

Proces tworzenia modelu nazywa się *modelowaniem* obiektu. Poszczególne jego etapy przedstawia Rys. 5.6. Pierwszym etapem modelowania wybranego procesu (fizjologicznego lub patologicznego) musi być zebranie dotychczasowej wiedzy o tym zjawisku. Wiedza ta może być reprezentowana zarówno przez modele werbalne, graficzne jak i w formie dotychczas istniejących modeli matematycznych.



Rys. 5.6 Etapy tworzenia modelu procesu chorobowego

Bardzo ważnym etapem modelowania jest określenie celu modelu. Można wyróżnić dwa podstawowe cele modelowania. Pierwszym przeznaczeniem modelu jest uzyskanie precyzyjnego opisu badanego obiektu. Drugim celem tworzenia modeli jest ich zastosowanie do problemów decyzyjnych (wspomaganie decyzji diagnostycznych, terapeutycznych). Oba te cele mogą determinować różną strukturę oraz różną postać wektora obserwacji dla tworzonego modelu.

Kolejnym etapem modelowania jest formułowanie modelu matematycznego, a więc określenie wektora obserwacji, stosownej struktury modelu oraz jego identyfikacja (czyli oszacowanie wartości parametrów).

Mając gotowy model matematyczny, należy go zweryfikować. Procedura weryfikacji modelu najczęściej obejmuje dwa etapy. W pierwszym z nich testuje się dopasowanie modelu do obserwacji. Innymi słowy sprawdza się, czy model odtwarza modelowany proces z założoną dokładnością. Pozytywny wynik tego testu pozwala przejść do weryfikacji merytorycznej modelu. Wówczas ocenia się zgodność modelu z przyjętymi założeniami oraz celem modelowania. Ponadto na tym etapie modelowania dokonuje się biomedycznej interpretacji wniosków wynikających z modelu matematycznego.

## **5.4 Klasyfikacja modeli**

W literaturze poświęconej teorii modelowania można znaleźć różne klasyfikacje modeli matematycznych, zależnie od przyjętego kryterium. W skrypcie tym ograniczono się do dwóch najważniejszych kryteriów: złożoność modelu oraz cel modelowania.

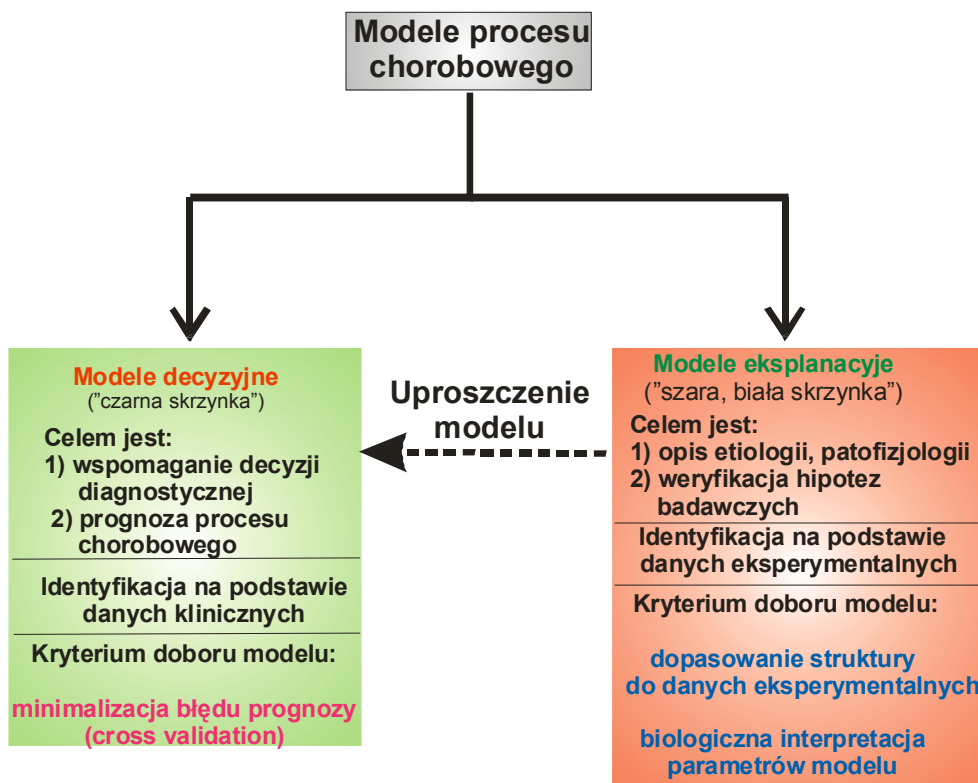
### **5.4.1 Klasyfikacja modeli ze względu na złożoność modelu**

W zależności od przeznaczenia danego modelu oraz posiadanej wiedzy o obiekcie, możemy budować modele o różnym poziomie szczegółowości. Załóżmy, że przedmiotem modelowania jest określenie związku między czynnikiem  $r(t)$  stymulującym dany proces a jego surogatem  $d(t)$ , bez uwzględnienia opisu przyczynowo-skutkowego. Ta klasa modeli nazywana jest modelami typu „czarnej skrzynki”. Jeżeli dysponujemy częściową wiedzą, bądź hipotezami dotyczącymi związków przyczynowo-skutkowych, otrzymujemy modele typu „szarej skrzynki”. Modele takie wyjaśniają tylko częściowo mechanizmy leżące u podstaw modelowanego procesu. Jeżeli celem modelowania jest zidentyfikowanie związków przyczynowo-skutkowych między obserwacjami, wówczas szukamy modelu typu „białej skrzynki”. Modele takie, najczęściej są budowane w oparciu na prawach fizyko-chemicznych i dostarczają największej wiedzy o modelowanym procesie. Często jednak zawierają dużo parametrów, które mogą być trudne do identyfikacji na podstawie obserwacji. Dlatego modele tego typu są rzadko budowane.

### **5.4.2 Klasyfikacja modeli ze względu na cel modelu**

Klasyfikację modeli uwzględniających cel modelowania przedstawia rys. 5.7. Można wyróżnić dwie klasy modeli: modele decyzyjne oraz modele eksplanacyjne.

*Modele decyzyjne* są konstruowane do wspomagania podejmowania decyzji diagnostycznych lub leczniczych. Na podstawie wektora obserwacji, złożonego z określonych cech klinicznych pacjenta model prognozuje (przewiduje) decyzję diagnostyczną. Z punktu widzenia teorii decyzji zagadnienie rozpoznania choroby jest równoważne zagadnieniu dyskryminacji, zaś problem diagnostycznego różnicowania odpowiada zagadnieniu klasyfikacji. Dlatego też modele decyzyjne najczęściej są typu „czarnej skrzynki”, gdzie zależności przyczynowo-skutkowe są modelowane *implicite*. Innym przykładem są modele pozwalające prognozować przyszły stan chorego, bądź odpowiedź na zastosowane leczenie na podstawie aktualnego stanu chorego. Modele diagnostyczne często są stosowane w radiologii, np. podczas przesiewowej mammografii. Tradycyjnie stosuje się wtedy ocenę mammogramu przez dwóch niezależnych radiologów. W niektórych ośrodkach klinicznych w USA ocenę pierwszego radiologa zastępuje się prognozą z modelu diagnostycznego.



Rys. 5.7 Klasyfikacja modeli ze względu na cel modelowania

Innym przykładem modeli decyzyjnych jest ich użycie do prognozy wystąpienia raka sutka lub raka jajnika na podstawie wyników badań genetycznych, wybranych cech klinicznych oraz występowania obciążeń rodzinnych.



Charakterystyczną cechą modeli decyzyjnych jest wymaganie, by parametry modelu mogły być zidentyfikowane na podstawie obserwacji dostępnych z badania klinicznego. Warunek ten można nazwać postulatem *klinicznej identyfikowalności modelu decyzyjnego*. Pożądanym kryterium wyboru „najlepszego” modelu decyzyjnego jest otrzymanie możliwie najlepszej prognozy poprawnych decyzji.

Drugą klasą modeli uwzględniającą cel modelowania są *modele eksplanacyjne*. Przeznaczeniem tej klasy modeli jest znalezienie opisu (wyjaśnienia mechanizmu) badanego zjawiska. Dlatego modele eksplanacyjne są najczęściej modelami typu szarej lub „białej skrzynki”. Mogą być one konstruowane po to, by zweryfikować hipotetyczny mechanizm zjawiska lub zidentyfikować zależności między obserwacjami, w przypadku braku hipotez dotyczących związków przyczynowo-skutkowych między badanymi czynnikami. Ponieważ modele eksplanacyjne często są tworzone na podstawie doświadczeń, warunek klinicznej identyfikowalności modelu nie jest pożądanym. Jednak budując model eksplanacyjny powinno się dążyć do tego, by parametry modelu, bądź jego elementy składowe posiadały interpretacje biologiczne. Przykładem takich modeli są modele kompartmentowe, w których parametry opisują stałe czasowe eliminacji substancji z kompartmentu (np. osocza krwi). Postulat biologicznej interpretacji elementów składowych modelu eksplanacyjnego, może wymuszać konieczność zbudowania mniej dokładniejszego modelu, ale za to dostarczającego więcej wiedzy o badanym obiekcie.

Niespełnienie warunku identyfikowalności klinicznej modelu eksplanacyjnego sprawia, że rzadko kiedy można go stosować jako model decyzyjny. Dlatego często stosuje się upraszczanie złożonych modeli eksplanacyjnych (np. przyjmując niektóre parametry jako niezmiennie międzyosobniczo, lecz identyfikowalne klinicznie) otrzymując w ten sposób modele decyzyjne. W ten sposób powstał na przykład model używany do szacowania insulinooporności na podstawie testu dożylnego (lub doustnego) obciążenia glukozą.

Przedstawiona klasyfikacja modeli może być pomocna przy wyborze właściwego dla danego problemu typu modelu. Zawsze jednak należy konstruować możliwie najprostszy model, który spełnia przyjęte cele modelowania. Dlatego wybór klasy i struktury modelu jest elementem sztuki, a nie techniki modelowania, która wymaga intuicji i naukowej mądrości. Można także powiedzieć, że mądrość (rozumiana w sensie potocznym) to umiejętność znajdowania kompromisu między niepotrzebną komplikacją a zbytnim upraszczaniem w postrzeganiu rzeczywistości.

## 6. PODSTAWY WNIOSKOWANIA W NAUKACH MEDYCZNYCH

### 6.1 Rodzaje wnioskowania naukowego

Konstrukcja modelu badanego zjawiska lub procesu chorobowego jest pierwszym etapem tworzenia wiedzy medycznej. Model matematyczny pozwala bowiem na agregację obserwacji oraz ekstrakcję zawartych w nich informacji. Przekształcenie tych informacji w wiedzę medyczną wymaga przeprowadzenia specyficznego procesu myślowego, zwanego *wnioskowaniem*. Wnioskowanie polega na formułowaniu nowych twierdzeń na podstawie przyjętych założeń oraz nowych informacji dostarczonych przez model. Założenia (najczęściej są one „wbudowane” w model) oraz informacje pochodzące z obserwacji tworzą *zbiór przesłanek*, zaś twierdzenia powstałe w wyniku wnioskowania nazywane są *wnioskami* lub *konkluzjami*.

W rozumowaniu naukowym wyróżnia się dwa główne sposoby wnioskowania. Pierwszy z nich, nazywane *dedukcją*, polega na wyprowadzeniu z pewnych zdań (prawdziwych przesłanek) wynikającego z nich logicznie następstwa (prawdziwego wniosku). Dedukcja pozwala formułować wnioski dotyczące szczególnych sytuacji na podstawie udowodnionych praw ogólnych. Przykładem rozumowania dedukcyjnego, stosowanym w praktyce klinicznej jest wnioskowanie o skuteczności proponowanego leczenia u danego pacjenta na podstawie wyników prób klinicznych zastosowanego leku. Lekarz wówczas *implicite* posługuje się następującym schematem myślenia (dla celów dydaktycznych opis jego został uproszczony). Ponieważ w próbach klinicznych wykazano, że lek *A* jest skuteczny w leczeniu choroby *B*, można wnioskować że lek ten będzie także skuteczny u konkretnego pacjenta chorego na chorobę *B*.

Drugi sposób wnioskowania, komplementarny do dedukcji nazywany jest *indukcją*. Polega ona na przechodzeniu od szczegółu do ogółu, tzn. formułowaniu praw ogólnych na podstawie uwodnionych faktów szczegółowych. Przykładami zastosowania rozumowania indukcyjnego w medycynie jest wnioskowanie epidemiologiczne, w którym wnioskuje się o stanie zdrowia badanej populacji na podstawie badania pacjentów reprezentatywnych dla tej populacji.

Wnioskowanie dedukcyjne oraz indukcyjne prowadzi do jednoznacznych wniosków tylko wtedy, gdy posługujemy się przesłankami, które są pewne. Sytuacja taka najczęściej występuje w naukach teoretycznych, np. matematyce. W medycynie (lub ogólniej w naukach przyrodniczych) wnioskowanie odbywa się na podstawie modelu, który jest obciążony nie-

pewnością. Dlatego pomimo, że we wnioskowaniach posługujemy się formalnymi metodami rozumowania, formułowane wnioski nie zawsze mogą być traktowane jako wnioski pewne i ostateczne. W sytuacji, gdy wnioskowanie odbywa się na podstawie modelu matematycznego budowanego przy wykorzystaniu obserwacji, stosuje się pewną odmianę wnioskowania indukcyjnego nazywanym *wnioskowaniem statystycznym*.

## 6.2 Wnioskowanie statystyczne

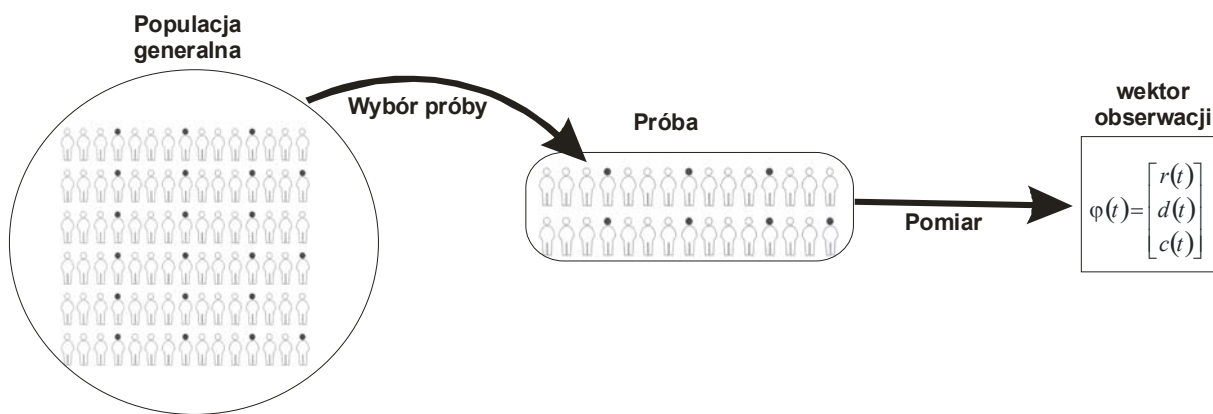
Wnioskowanie statystyczne polega na formułowaniu twierdzeń dotyczących badanego zjawiska na podstawie zarejestrowanych obserwacji. Załóżmy, że celem poznania naukowego jest określenie częstości występowania choroby w interesującej badacza populacji. Ze względów czasowych i kosztów ekonomicznych nie jest możliwe zbadanie wszystkich przedstawicieli danej populacji. Można jednak pokazać, że wystarczy zaobserwować jak często występuje dana choroba w pewnej grupie pacjentów należnych do analizowanej populacji i na tej podstawie wnioskować o częstości w całej populacji. Wnioskowanie to będzie oczywiście obciążone niepewnością, dlatego uzyskana wartość częstości będzie tylko pewnym oszacowaniem wartości rzeczywistej.

Wnioskowanie statystyczne obejmuje dwie grupy zagadnień. Pierwsze z nich, które zostało przedstawione w powyższym przykładzie, dotyczy problemu szacowania wybranego parametru zjawiska lub modelu, na podstawie wektora obserwacji. Zagadnienie to nazywa się *estymacją parametrów*.

Jednak oszacowanie wartości danego parametru nie zawsze pozwala na uzyskanie odpowiedzi na pytanie badawcze. Jeżeli badacza interesuje związek między częstością występowania choroby a rasą, estymacja wartości zachorowalności nie pozwoli rozstrzygnąć o istnieniu takiego związku. Wówczas posługujemy się drugą grupą metod wnioskowania statystycznego, zwaną *testowaniem hipotez*.

Elementem wspólnym dla estymacji parametrów i testowania hipotez są dwa elementy wnioskowania statystycznego: określenie populacji generalnej oraz próby losowej. *Populacją generalną* nazywa się skończony zbiór elementów, których będzie dotyczyło przeprowadzane wnioskowanie. W medycynie populacja najczęściej obejmuje zbiorowość ludzi lub zwierząt laboratoryjnych.

*Próbą* nazywa się wybrany podzbiór populacji generalnej, na którego elementach dokonuje się obserwacji badanych cech. Zależność między populacją, próbą i wektorem obserwacji przedstawia rys. 6.1



Rys. 6.8 Zależności między pojęciami „populacja generalna”, „próba” oraz „wektor obserwacji”

### 6.3 Wybór próby z populacji generalnej

Procedurę wyboru próby z populacji generalnej określa się mianem *próbkiowania* (ang. *sampling*). Właściwe wybieranie próby z danej populacji nie jest prostym zadaniem. W literaturze statystycznej i epidemiologicznej zagadnieniu temu poświęcone są oddzielne monografie. Dlatego w skrypcie przedstawione będą tylko podstawowe zasady pobierania próby z populacji.

Podstawowym warunkiem, który powinna spełniać wybrana próba jest jej *reprezentatywność*. Postuluje się wówczas fakt, że próba powinna odzwierciedlać w mikroskali istotne cechy badanej populacji. Innymi słowy, jeżeli badaną populację stanowią osoby w wieku 20-50 lat, to pobrana próba powinna mieć taki sam rozkład względem płci wieku, wykształcenia, miejsca zamieszkania itd., jaki istnieje w badanej populacji. Warunek ten często jest realizowany poprzez zapewnienie takiego samego prawdopodobieństwa wyboru każdego elementu populacji. Przykładowo, jeżeli w badanej populacji jest przewaga kobiet, to spełnienie tego warunku sprawi, że kobiety będą częściej wybierane do próby. Zatem rozkład płci w pobranej próbie będzie zbliżony do rozkładu, który byłby obserwowany w całej populacji.

Spełnienie warunku reprezentatywności próby prowadzi do tzw. *losowego pobierania próby*. Wówczas osoba konstruująca próbę nie ma żadnego wpływu na zakwalifikowanie danego elementu populacji do próby. Decyzję taką podejmuje „maszyna losująca” wybierająca elementy do próby z zadaniem schematem losowania. Załóżmy dla przykładu, że badaną populacją stanowią dorośli mieszkańcy Polski, którzy są niepełnosprawni fizycznie. Informacje o takiej populacji można otrzymać z GUS, który dysponuje bazą danych osób z orzeczoną niepełnością. Każdej osoby można przypisać numer porządkowy. Następnie, używając oprogramowania komputerowego generującego liczby losowe (o podobnym działaniu jak

maszyna losująca w grach losowych), można wylosować z bazy  $N$  osób. Zbiór tych osób będzie stanowił próbę losową badanej populacji. Zapewnienie losowego pobrania próby z danej populacji ma bardzo duże znaczenie przy kontroli czynników zakłócających, co zostanie omówione w następnym rozdziale.

Mając wylosowaną próbę można przejść do etapu wnioskowania statystycznego, czyli estymacji parametrów i testowania hipotez.

## 6.4 Estymacja parametrów

Zagadnienie estymacji parametrów zostanie omówione na przykładzie znajdowania wartości średniej cechy w wybranej populacji. Podstawy teoretyczne zadania estymacji wykraczają poza zakres metodologii poznania naukowego i zostaną omówione na kursie biostatystyki.

Założmy zatem, że celem prowadzonego badania jest wyznaczenie średniego wieku menopauzy u Polek. W tym celu wybrano losowo  $N$  kobiet w wieku 40-60 lat z terenu całej Polski, zarejestrowano wiek wystąpienia u nich ostatniej miesiączki. Otrzymane wartości wieku ostatniej miesiączki można zapisać w postaci zbioru liczb  $\{x_1, x_2, \dots, x_N\}$ , w którym  $x_i$  oznacza wiek wystąpienia menopauzy u  $i$ -tej kobiety. Wyznaczenie wieku średniego jest zadaniem prostym i sprowadza się do policzenia średniej arytmetycznej z liczb  $\{x_1, x_2, \dots, x_N\}$ , czyli:

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (4.8)$$

Powstaje jednak pytanie o interpretację otrzymanego wyniku. Czy wartość  $\bar{x}$  reprezentuje średni wiek menopauzy w badanej populacji czy tylko w próbie? Pomijając zbyteczne rozważania matematyczne, należy przyjąć, że wyznaczona średnia z próby jest bardzo dobrym oszacowaniem rzeczywistego średniego wieku menopauzy w badanej populacji. Jest jednak istotna różnica między średnim wiekiem w próbie i populacji. Dla próby  $\bar{x}$  reprezentuje rzeczywistą wartość średnią wieku menopauzy w próbie. Dla populacji -  $\bar{x}$  jest estymatorem wartości średniej wieku menopauzy. Mówiąc nieprecyzyjnie *estymatorem* pewnej wielkości będziemy nazywać funkcję określoną na wartościach próby (obserwacjach), pozwalającą na oszacowanie rzeczywistej (występującej w populacji) wartości badanego parametru. Zwyczajowo, estymator wartości średniej oznaczamy jako  $\hat{\bar{x}}$  (dodając znaczek „daszka” nad symbolem średniej). Zatem możliwe są dwa zapisy. Zapis  $\bar{x} = 50$  mówi, że wartość średnia

wieku menopauzy w próbie wyniosła 50 lat. Drugi zapis  $\hat{\bar{x}} = 50$  oznacza, że oszacowana wartość średnia dla badanej populacji wyniosła 50 lat.

Ponieważ każde oszacowanie dokonywane jest z pewnym błędem, prezentując oszacowaną wartość średnią dla populacji, należy zawsze podać wartość błędu estymacji (który ideologicznie odpowiada pojęciu błędu pomiaru). Można to zrobić na dwa sposoby. Pierwszy z nich jest następujący:

$$\hat{\bar{x}} = \bar{x} \pm SE \quad (4.9)$$

gdzie SE jest błędem standardowym będącym miarą błędu szacowania średniej na podstawie wylosowanej próby. Otrzymujemy w ten sposób *estymator punktowy* wartości średniej. Estymator ten jest najczęściej używany przy prezentacji wartości średniej w populacji.

Alternatywnym sposobem jest podanie przedziału ufności dla danego parametru. Przedział ufności można imprezować jako obraz zmienności uzyskiwanych wartości średnich w zależności od wylosowanych prób. Jeżeli byśmy pobierali 100 razy próbę z danej populacji i za każdym razem wyznaczali wartość średnią z próby, to otrzymalibyśmy co najwyżej 100 różnych wartości. Przedział ufności przedstawia zakres zmienności tych średnich.

Sposób wyznaczania błędu standardowego, jak i przedziału ufności zostanie przedstawiony na kursie biostatystyki. W najnowszym piśmiennictwie preferuje się estymacje przedziałowe, które dostarczają więcej informacji o dokładności oszacowania parametru.

## 6.5 Testowanie hipotez

Estymacja wartości parametrów w wybranej populacji nie zawsze musi być celem wnioskowania statystycznego. Załóżmy, że celem badania jest porównanie wartości średnich w dwóch populacjach. W tej sytuacji nie ma potrzeby znajomości wartości dwóch średnich. Wystarczający jest bowiem wniosek rozstrzygający zależność  $\hat{\bar{x}}_1 = \hat{\bar{x}}_2$ . Rozstrzygnięcie tej kwestii wymaga sformułowania hipotezy, która będzie zweryfikowana na podstawie wylosowanej próby. Weryfikacja hipotez polega na sprawdzaniu, czy są wystarczające przesłanki do odrzucenia założonej hipotezy na rzecz hipotezy alternatywnej. Mamy więc zawsze do czynienia z *hipotezą zerową* (weryfikowaną) oraz *hipotezą alternatywną* będącą negacją hipotezy zerowej. W rozważanym przykładzie hipotezą zerową jest przypuszczenie, że  $\hat{\bar{x}}_1 = \hat{\bar{x}}_2$ , zaś hipoteza alternatywna przyjmuje postać  $\hat{\bar{x}}_1 \neq \hat{\bar{x}}_2$ . Do podjęcia decyzji czy zebrane obserwacje pozwalają na odrzucenie hipotezy zerowej służą *statystyki testowe*. Wyznaczenie wartości statystyki testowej na podstawie wartości występujących w wybranej próbie umożliwia podjęcie właściwej decyzji. Co więcej, statystyki testowe są w ten sposób konstruowane, aby

decyzja podjęta na podstawie dwóch prób, uogólniała się na decyzję dotyczącą dwóch populacji. Należy zawsze pamiętać, że nie odrzucenie hipotezy zerowej nie dowodzi jej prawdziwości. Test statystyczny wykazuje jedynie brak podstaw do jej odrzucenia. Dlatego jako hipotezę zerową przyjmuje się zawsze tę, dla której szukamy przesłanek do odrzucenia.

Weryfikacja hipotezy zerowej prowadzi do wyznaczenia tzw. *prawdopodobieństwa testowego*. Wyraża ono prawdopodobieństwo odrzucenia prawdziwej hipotezy zerowej, a zatem jest miarą błędu wnioskowania. Im mniejsza jest jego wartość, tym większe jest nasze przekonanie o słuszności odrzucenia hipotezy zerowej.

Szczegółowy opis procedury testowania hipotez statystycznych wykracza poza zakres tematyczny tego skryptu. Jest on prezentowany w ramach kursu z biostatystyki.

## 6.6 Ocena zależności między obserwacjami

Kombinacja omówionych procedur wnioskowania statystycznego jest stosowana przy ocenie zależności między obserwacjami. Załóżmy, że celem eksperymentu naukowego jest budowa modelu opisującego wpływ masy ciała kobiety na jej płodność. Po zdefiniowaniu populacji badanej oraz wybraniu odpowiedniej próby dokonujemy obserwacji masy ciała kobiety oraz pomiaru stężenia progesteronu w drugiej fazie cyklu miesięczkowego (będącego wykładnikiem jajeczkowania). Pierwszym krokiem w budowaniu modelu powinno być zbadanie, czy istnieje jakaś zależność między masą ciała a stężeniem progesteronu. Brak takiej zależności wykaże bowiem bezzasadność budowy takiego modelu.

Do oceny istnienia zależności między obserwacjami służą *współczynniki korelacji*. Są to również pewne wielkości, których wartości są szacowane na podstawie próby. Wyznaczenie wartości współczynnika korelacji pozwala nam wnioskować o sile i kierunku badanych zależności. Potencjalne wartości współczynników korelacji zawierają się w przedziale  $[-1;1]$ . Wartość dodatnia świadczy o korelacji dodatniej oznaczający ten sam kierunek zmian dwóch obserwacji. Wartość ujemna oznacza sytuację przeciwną. Im obliczona wartość współczynnika korelacji jest bliższa 1 lub  $-1$ , tym związek między obserwacjami jest silniejszy. Jeżeli w wyniki obliczeń otrzymuje się wartość zero, oznacza to brak zależności między zmiennymi.

Obliczenie niezerowej wartości współczynnika korelacji na podstawie obserwacji w próbie nie upoważnia do wnioskowania, że istnieje związek w całej badanej populacji. Takie wnioskowanie uogólniające jest możliwe tylko wtedy, gdy w wyniku procedury testowania hipotez, zostanie odrzucona hipoteza o zerowej wartości współczynnika korelacji. Dlatego przy wnioskowaniu statystycznym dotyczącym oceny zależności między obserwacjami jest

niezbędna zarówno wartość współczynnika korelacji, jak też wartość prawdopodobieństwa testowego.

## 6.7 Pułapki wnioskowania statystycznego

Podczas edukacji szkolnej i akademickiej jesteśmy uczeni sprawnego posługiwania się wnioskowaniem dedukcyjnym i indukcyjnym w odniesieniu do rozwiązań problemów z przedmiotów teoretycznych. Bagaż tych umiejętności jest tak „ciężki”, że często zapominamy o specyfice wnioskowania statystycznego i stosujemy w nim klasyczne zasady indukcji. Takie podejście prowadzi do następujących pułapek. Po pierwsze wnioski z wnioskowania statystycznego są wówczas formułowane jako wnioski „ostateczne”, pewne, nie budzące żadnych wątpliwości. Takie podejście jest sprzeczne z przedstawioną ideologią wnioskowania statystycznego, które *per se* zawiera element niepewności.

Drugą, często spotykaną pułapką wnioskowania statystycznego jest ekstrapolacja wniosków poza badaną populację. Takie rozumowanie jest niedopuszczalne, gdyż poprawnie przeprowadzone wnioskowanie statystyczne może dostarczyć wiedzy tylko o populacji, z której była wybrana próba. Ekstrapolacja wniosków na szerszą populację w stosunku do populacji badanej wymaga przeprowadzenia powtórnego wnioskowania na podstawie próby pobranej z rozszerzonej populacji. Zagadnienie nieuzasadnionej ekstrapolacji jest często powodem wielu konfuzji w środowiskach naukowych. Jednym z problemów z tym związanych jest dyskusja nad przenoszeniem wniosków formułowanych na podstawie badań na zwierzętach na wnioski dotyczące populacji ludzkiej. W tej sytuacji mamy także do czynienia z rozszerzaniem badanej populacji, a zasadnicze pytanie brzmi, czy populacja zwierząt laboratoryjnych jest podobna do populacji ludzkiej? Im większe jest to podobieństwo (w odniesieniu do analizowanego procesu biologicznego), tym bardziej możliwa jest ekstrapolacja wyciąganych wniosków.

Trzecim źródłem pułapki związanej z wnioskowaniem statystycznym jest liczebność pobieranej próby. Im jest ona mniejsza, tym większa jest niepewność wyciąganych wniosków. Co więcej, wnioski wyciągane na podstawie zbyt małej próby mogą być całkowicie odmienne jakościowo od wniosków formułowanych na podstawie wystarczająco licznej próby. Rozwój technik komputerowych i pakietów statystycznych sprawia, że dla większości stosowanych testów statystycznych można wyznaczyć minimalną liczebność próby warunkującą poprawne wnioskowanie statystyczne.



Czwarta pułapka wnioskowania statystycznego związana jest z utożsamianiem korelacji między obserwacjami ze związkiem przyczynowo – skutkowym. Problemowi temu będzie poświęcony następny rozdział.

Literatura uzupełniająca;

1. K. Szamotulska, Z. Brzeziński: *Epidemiologia kliniczna* .PZWL 1997
2. W. Wędrychowski *Epidemiologia. Wprowadzenie i metody badań* 1999
3. J. Koronacki, J. Mielniczuk *Statystyka dla kierunków technicznych i przyrodniczych*. WNT 2001
4. A. Stanisław *Biostatystyka*. Kraków 2005
5. Jerzy Apanowicz *Metodologia nauk*, "Dom organizatora", Toruń 2003.