

Materiały pomocnicze do laboratorium

Biometria i nadzór wizyjny



Politechnika Poznańska
©Sławomir Maćkowiak

Materiały są przeznaczone wyłącznie do użytku podczas zajęć z przedmiotu *Biometria i nadzór wizyjny* prowadzonych przez dr. Inż. Sławomir Maćkowiaka oraz dr. Inż. Damiana Karwowskiego na studiach na kierunku Elektronika i Telekomunikacja.

Uwaga: Niektóre ilustracje pochodzą z podręczników zalecanych jako pomocnicze do przedmiotu.

Kopiowanie, rozpowszechnianie i używanie w innych celach jest stanowczo zabronione.

W szczególności, nieuprawnione rozpowszechnianie drogą elektroniczną np. w Internecie jest nielegalne

SYSTEM BERTILLON'EGO – STATYSTYCZNA ANALIZA DANYCH

Ćwiczenie dotyczy statycznych pomiarów z zakresu antropometrii w odniesieniu do zastosowań identyfikacji biometrycznej zaproponowanej przez Bertillon'ego. W ćwiczeniu student zapoznaje się z punktami antropometrycznymi, przyrządami pomiarowymi i techniką pomiaru. Poszczególne tematy zadań dotyczą doboru cech pomiarowych, komponowania karty badań. Główny nacisk ćwiczenia położony jest na metody opracowywania danych pomiarowych.

1. Wprowadzenie – pomiary Bertillon'ego

1.1 Antropometria

Nazwa „antropometria” wywodzi się z greckich słów anthropos - człowiek, metrem - miara. Zadaniem antropometrii jest przełożenie rozmiarów i kształtów ciała człowieka na liczby i określone stosunki ilościowe. Wyróżnia się antropometrię statyczną i dynamiczną (goniometrię).

a) Antropometria statyczna. Zajmuje się pomiarem cech w pozycjach nieruchomych. Dane antropometryczne określają wymiary i proporcje między podstawowymi odcinkami ciała człowieka. Pomiary antropometryczne wykonywane są zgodnie z przyjętą techniką pomiarową w/g Międzynarodowej Organizacji Standaryzacji. Osoby stojące przyjmują podczas pomiarów pozycję zasadniczą, natomiast niemowlęta i osoby leżące mierzy się w pozycji leżącej. Cechy długościowe ciała są mierzone antropometrem, cechy szerokościowe cyrklem kabłąkowym, a obwody ciała centymetrem krawieckim. Przy pomiarach głowy pacjent siedzi tak, aby umożliwić dostęp do głowy ze wszystkich stron.

Układ odcinków ciała w swobodnej pozycji wyprostowanej tworzy postawę ciała człowieka. Mianem postawy prawidłowej określa się harmonijne ułożenie poszczególnych odcinków ciała względem siebie oraz w odniesieniu do głównej osi ciała, gdy napięcie układu mięśniowego i nerwowego jest minimalne.

Dokonywane pomiary można podzielić na:

- pomiary wysokości (wysokość ciała, długość kończyn dolnych itd.),
- pomiary długości (ramienia, tułowia, kończyny górnej itd.),
- pomiary szerokości i głębokości (barków, bioder itd.),
- pomiary obwodów (głowy, szyi, pasa itd.),
- inne (średnica chwytu ręki itd.).

Pomiary antropometryczne statyczne są wykonywane względem stałych punktów antropometrycznych.

b) Goniometria. Dział antropometrii (gonion - kąt, metron - miara) zajmujący się pomiarami kątów między różnymi odcinkami ciała ludzkiego lub kości. Wyróżnia się:

- goniometrię statyczną (pomiar wielkości kątów między płaszczyznami i odcinkami ciała lub kości),

- goniometrię dynamiczną (pomiar zakresów ruchu w poszczególnych stawach organizmu - tułowia, kończyn górnych i dolnych).

Pomiary te określają możliwości zasięgów oraz rozpiętości ruchów.

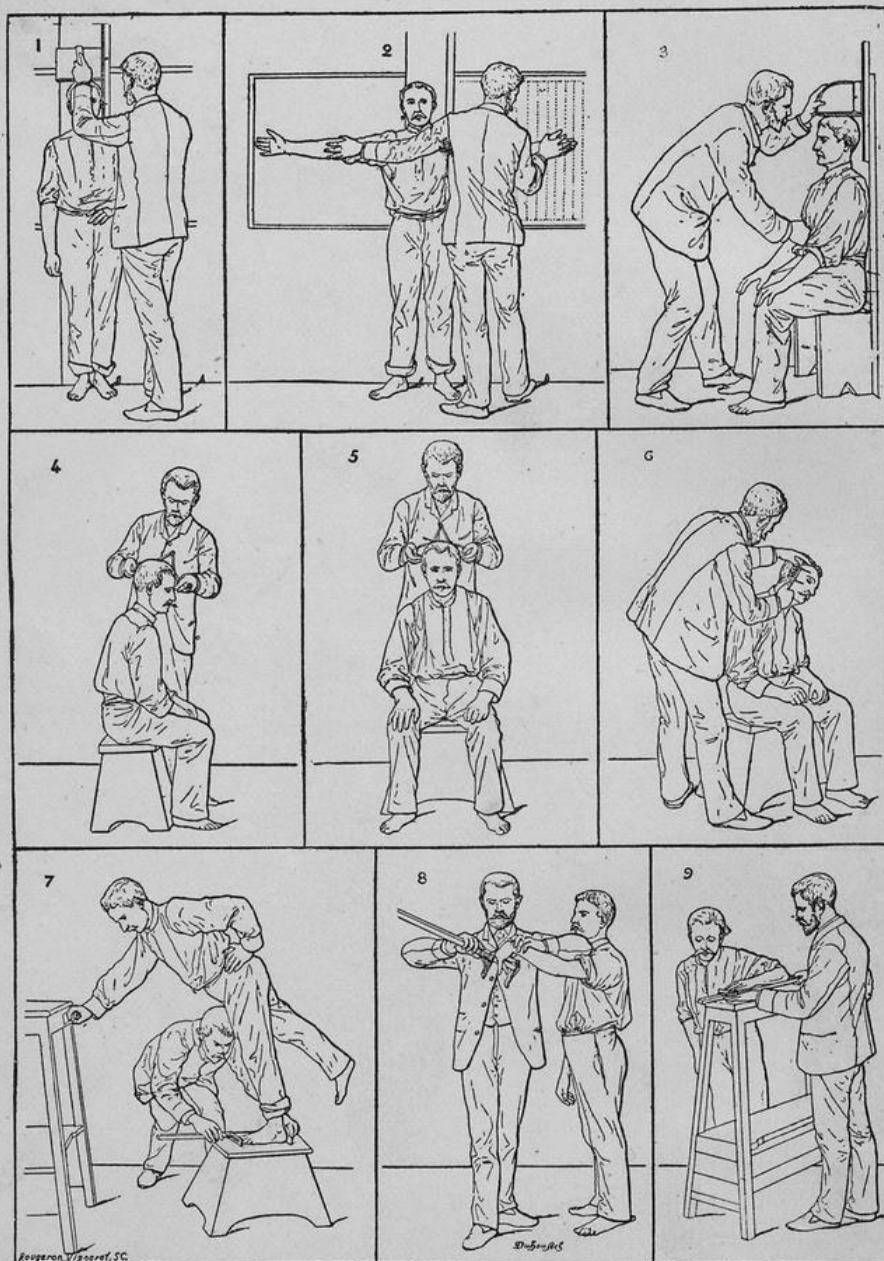
1.2 Pomiary Bertillon'ego

Bertillon took measurements of certain bony portions of the body, among them the skull width, foot length, cubit, trunk and left middle finger. These measurements, along with hair color, eye color and front and side view photographs, were recorded on cardboard forms measuring six and a half inches tall by five and a half inches wide.

By dividing each of the measurements into small, medium and large groupings, Bertillon could place the dimensions of any single person into one of 243 distinct categories. Further subdivision by eye and hair color provided for 1,701 separate groupings.

Upon arrest, a criminal was measured, described and photographed. The completed card was indexed and placed in the appropriate category. In a file of 5,000 records, for example, each of the primary categories would hold only about 20 cards. It was therefore not difficult to compare the new record to each of the other cards in the same category. If a match was discovered, the new offense was recorded on the criminal's card.

**ABSTRACT OF
THE ANTHROPOMETRICAL SIGNALMENT**



1. Height.

4. Length of head.

7. Left foot.

2. Reach.

5. Width of head.

8. Left middle finger.
(ii)

3. Trunk.

6. Right ear.

9. Left forearm.

2. Analiza statystyczna

Zmienna losowa

Zmienną (cechę) X nazywamy losową (przypadkową) jeżeli nie jesteśmy w stanie przewidzieć jej wartości, gdyż wartość jej zależy od wyniku losowania (od przypadku). Przez zmienną losową można zatem rozumieć wielkość, która w wyniku doświadczenia przyjąć może wartość z pewnego zbioru liczb i to z określonym prawdopodobieństwem.

Zmienne losowe oznaczamy zazwyczaj dużymi literami alfabetu łacińskiego np. X, Y, Z . Wartości przybierane przez zmienne losowe (zwane realizacjami) oznaczane są odpowiednimi małymi literami np. x, y, z .

Zdarzenie losowe: zmienna losowa X przyjmuje wartość x_0 , oznaczamy wyrażeniem $(X=x_0)$. Miarą szans zajścia zdarzenia losowego $(X=x_0)$ jest jego prawdopodobieństwo $P(X=x_0)$ będące liczbą z przedziału $[0, 1]$. Prawdopodobieństwo zdarzenia pewnego jest równe 1 , a prawdopodobieństwo zdarzenia niemożliwego jest równe 0 . Prawdopodobieństwo, że zdarzenie $(X=x_0)$ nie zajdzie, jest równe $1-P(X=x_0)$.

Wyróżniamy dwa typy zmiennych losowych:

- **zmienna losowa skokowa** (dyskretna), o skończonej lub przeliczalnej liczbie możliwych wartości (realizacji)
- **zmienna losowa ciągła**, mogąca przyjmować dowolne wartości z określonego przedziału liczb rzeczywistych

Każda zmienna losowa jest określona przez swój rozkład prawdopodobieństwa (prawdopodobieństwa z jakim przyjmuje ona poszczególne wartości).

W celu sporządzenia rozkładu prawdopodobieństwa zmiennej losowej skokowej X podaje się zbiór jej możliwych wartości oraz funkcję $p(x_i) = P(X = x_i)$, określającą prawdopodobieństwo zdarzenia losowego $X = x_i$ dla każdej możliwej wartości x_i ($i = 1, 2, \dots$) danej zmiennej losowej X , nazywaną funkcją prawdopodobieństwa zmiennej losowej X .

Zachodzi przy tym zależność $\sum_{i=1}^{\infty} p(x_i) = 1$ dla $i = 1, 2, \dots$

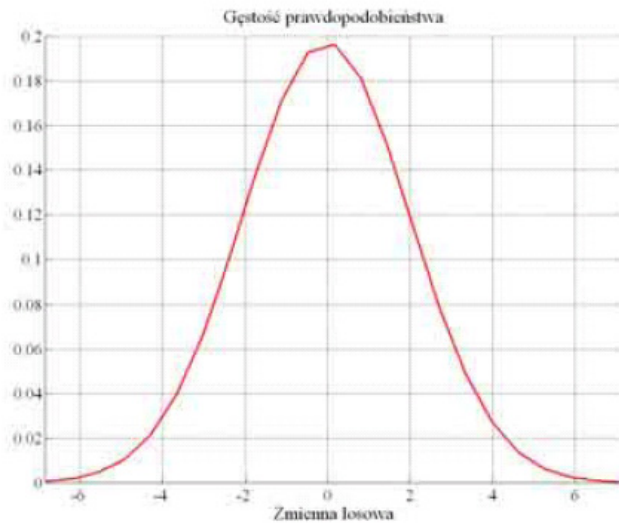
W celu sporządzenia rozkładu prawdopodobieństwa zmiennej losowej ciągłej X konieczna jest znajomość nieujemnej funkcji $f(x)$ zwanej funkcją gęstości prawdopodobieństwa zmiennej losowej X albo krócej funkcją gęstości lub wprost gęstością zmiennej losowej X .

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x}$$

Prawdopodobieństwo zdarzenia losowego, że zmienna losowa X typu ciągłego przyjmuje wartość z przedziału (a, b) – skończonego lub nieskończonego – jest całką funkcji gęstości prawdopodobieństwa w tym przedziale:

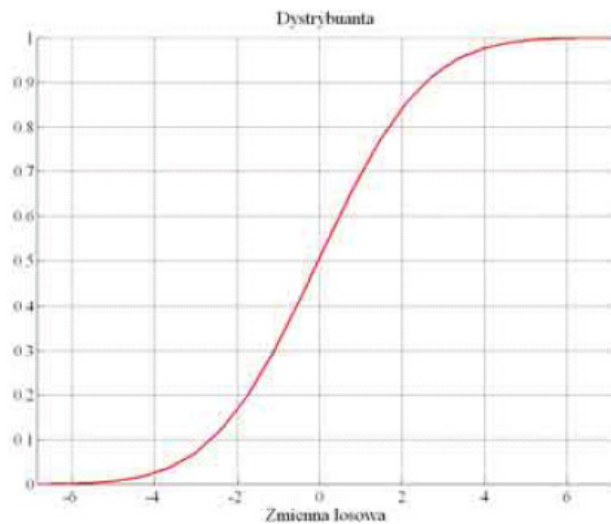
$$P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx$$

Zachodzi przy tym zależność: $\int_{-\infty}^{+\infty} f(x) dx = 1, f(x) \geq 0$



Ważnym pojęciem związanym ze zmienną losową i jej rozkładem prawdopodobieństwa jest pojęcie dystrybuanty (prawdopodobieństwo zdarzenia losowego: zmienna losowa X przyjmuje wartości mniejsze od liczby x):

$$F(x) = P(X < x)$$



Własności dystrybuanty:

- $0 \leq F(x) \leq 1$
- dla $x_1 < x_2$ zawsze $F(x_1) \leq F(x_2) \rightarrow$ funkcja niemalejąca
- $P(a \leq X < b) = F(b) - F(a)$

Szereg rozdzielczy

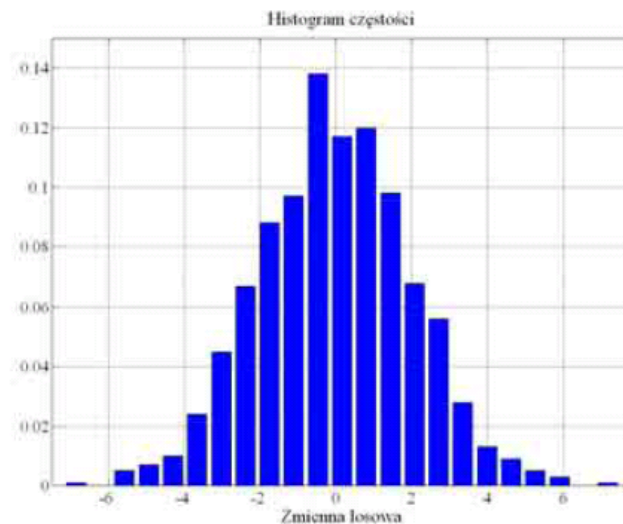
Klasyfikacja danych statystycznych to procedura uporządkowania danych, polegająca na podziale zbioru danych na przedziały zwane klasami. Każdy element zbioru danych może być zaliczony tylko do jednej klasy.

Liczbę n_i danych, które należą do i -tej klasy nazywamy jej liczebnością. Stosunek liczebności n_i i -tej klasy do liczebności N wszystkich danych nazywamy częstością i -tej klasy $c_i = n_i/N$.

Szeregiem rozdzielczym liczebności (empirycznym rozkładem zmiennej losowej X) nazywamy zbiór uporządkowanych par (środek klasy x_i , liczebność klasy n_i), w których każdej klasie przyporządkowana jest jej liczebność $\left(\sum_{i=1}^k n_i = N\right)$.

Szeregiem rozdzielczym częstości (empirycznym rozkładem prawdopodobieństwa zmiennej losowej X) nazywamy zbiór uporządkowanych par (środek klasy x_i , częstość klasy c_i), w których każdej klasie przyporządkowana jest jej częstość $\left(\sum_{i=1}^k c_i = 1\right)$.

Wykres szeregu rozdzielczego nazywamy histogramem (liczebności, częstości).



Rozkład normalny (Gaussa)

Rozkład, którego gęstość prawdopodobieństwa wyraża się wzorem:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

gdzie: μ - wartość oczekiwana (średnia),

σ - odchylenie standardowe,

nazywany rozkładem normalnym i oznaczamy symbolem $N(\mu, \sigma)$.

Własności rozkładu normalnego:

- $P(|X-\mu| < \sigma) \approx 0,682$
- $P(|X-\mu| < 2\sigma) \approx 0,954$
- $P(|X-\mu| < 3\sigma) \approx 0,998$

W przyrodzie i technice spotykamy często rozkłady „normalne” w tym sensie, że są bardzo zbliżone do rozkładu normalnego. Wszędzie tam, gdzie wartość liczbową danej cechy

kształtuje się pod wpływem dużej liczby czynników i żaden z nich nie góruje nad pozostałymi, można spodziewać się występowania rozkładu normalnego. Przykładowo: przypadkowe błędy pomiarów, błędy obróbki mechanicznej, cechy elementów pochodzących z masowej produkcji mają na ogół rozkłady normalne.

Estymacja przedziałowa

Estymator to wielkość wyznaczona na podstawie próby losowej, służąca do oceny wartości nieznanymi parametrów populacji generalnej. Estymator, jako wielkość obliczana na podstawie próby losowej, jest zmienną losową. Błędem szacunku (estymacji) parametru populacji generalnej nazywamy różnicę między uzyskaną liczbową oceną parametru i jego prawdziwą wartością.

Estymacja przedziałowa polega na wyznaczeniu przedziału liczbowego, który z określonym prawdopodobieństwem β będzie zawierał nieznaną wartość szacowanego parametru. Przedział taki nazywamy przedziałem ufności, a prawdopodobieństwo β – poziomem ufności.

Wybór poziomu ufności jest arbitralny, ale od jego wartości zależy długość przedziału ufności. Dla prób o ustalonej liczności i w przypadku ustalonego estymatora im wyższy poziom ufności, tym większa długość przedziału ufności, ale tym mniejsze ryzyko nieobjęcia przez przedział ufności prawdziwej wartości szacowanego parametru.

Charakterystyki rozkładu zmiennej losowej

Do charakterystyk najczęściej wykorzystywanych w analizie właściwości rozkładu zmiennej losowej należą:

- **miary średnie (położenia)** – służące do określania tej wartości zmiennej losowej, wokół której skupiają się wszystkie pozostałe wartości zmiennej,
- **miary rozproszenia** – służące do badania stopnia zróżnicowania wartości zmiennej losowej,
- **miary asymetrii (skośności)** – służące do badania kierunku zróżnicowania wartości zmiennej losowej,
- **miary koncentracji** – służące do analizy stopnia skupienia wartości zmiennej losowej wokół średniej.

Przykładem miary średniej jest wartość średnia (średnia arytmetyczna):

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^N x_i$$

Przykładem miary rozproszenia jest odchylenie standardowe:

$$s_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$$

Przykładem miary asymetrii jest współczynnik asymetrii (skośności):

$$A_x = \frac{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2} \right)^3}$$

Gdy rozkład jest symetryczny to $A_x=0$, w przypadku asymetrii lewostronnej $A_x<0$, a w przypadku asymetrii prawostronnej $A_x>0$.

Przykładem miary koncentracji jest kurtoza:

$$K_x = \frac{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2}$$

Dowodzi się, że dla rozkładu normalnego $K_x=3$, dla spłaszczonego $K_x<3$, a dla wysmukłego $K_x>3$.

Analiza statystyczna w pakiecie Matlab

Generowanie próby losowej o rozkładzie normalnym

Składnia funkcji:

```
probka = e_normrnd(wo,sigma,m,n);
probka = normrnd(wo,sigma,m,n);
```

Wyniki:

probka – macierz prób losowych o m wierszach i n kolumnach.

Parametry:

wo – wartość oczekiwana (średnia) rozkładu
sigma – odchylenie standardowe rozkładu
m – liczba wierszy macierzy wynikowej
n – liczba kolumn macierzy wynikowej

Przykład zastosowania:

```
probka = e_normrnd(0,2,1,1000);
```

Identyfikacja rozkładu normalnego

Składnia funkcji:

```
e_normplot(probka)
normplot(probka)
```

Wyniki:

Graficzny test normalności rozkładu danych zawartych w próbie losowej *probka*.
Jeśli dane pochodzą z rozkładu normalnego, to wykres jest liniowy. Jeśli dane pochodzą z innego rozkładu, to wykres traci charakter liniowy.

Parametry:

probka – testowana próba losowa

Przykład zastosowania:

```
figure  
e_normplot(probka)  
whitebg(gcf)
```

Estymacja parametrów rozkładu normalnego

Składnia funkcji:

```
[wo_e,sigma_e,wo_uf,sigma_uf] = e_normfit(probka,1-puf)  
[wo_e,sigma_e,wo_uf,sigma_uf] = normfit(probka,1-puf)
```

Wyniki:

wo_e – estymowana wartość średnia
sigma_e – estymowane odchylenie standardowe
wo_uf – przedział ufności wartości średniej
sigma_uf – przedział ufności odchylenia standardowego

Parametry:

probka – wektor próby losowej
puf – poziom ufności

Przykład zastosowania:

```
puf = 0.95; % poziom ufności  
[wo_e,sigma_e,wo_uf,sigma_uf] = e_normfit(probka,1-puf)
```

Obliczanie charakterystyk rozkładu zmiennej losowej

Składnia funkcji (współczynnik asymetrii):

```
a = e_skewness(probka)  
a = skewness(probka)
```

Wyniki:

a – współczynnik asymetrii (skośności) danych zawartych w próbie losowej *probka*

Parametry:

probka – wektor próby losowej

Składnia funkcji (kurtoza):

```
k = e_kurtosis(probka)  
k = kurtosis(probka)
```

Wyniki:

k – kurtoza danych zawartych w próbie losowej *probka*

Parametry:

probka – wektor próby losowej

Histogram liczebności

Składnia funkcji:

```
[liczebnosci,dziedzina] = hist(probka,licz_klas);
```

Wyniki:

liczebnosci – wektor liczebności poszczególnych klas
dziedzina – dziedzina histogramu

Parametry:

probka – wektor próby losowej
licz_klas – założona liczba klas

Przykład zastosowania:

```
[liczebnosci,dziedzina] = hist(probka,20);  
figure  
bar(dziedzina,liczebnosci)  
grid
```

Estymacja gęstości zmiennej losowej

Składnia funkcji:

```
[gesto,dziedzina] = e_ksdensity(probka);  
[gesto,dziedzina] = ksdensity(probka);
```

Wyniki:

gesto – funkcja gęstości prawdopodobieństwa danych zawartych w próbie losowej
probka
dziedzina – dziedzina funkcji gęstości prawdopodobieństwa

Parametry:

probka – wektor próby losowej

Przykład zastosowania:

```
[gesto,dziedzina] = e_ksdensity(probka);  
figure  
plot(dziedzina,gesto)  
grid
```

3. Prosty klasyfikujący system bazo-danowy

W ogólności każda baza danych jest złożona z elementów o określonej strukturze - rekordów, czyli podstawowej jednostki informacji zawierającej pełny zestaw informacji o gromadzonych obiektach (w naszym przypadku będą to dane antropometryczne osób).

Za sterowanie danymi w bazie odpowiedzialny jest system zarządzania bazą danych (database management system – DBMS). DBMS może m. in. : formatować raporty, importować i eksportować dane, a także dzięki własnemu językowi skryptów formułować zapytania (query).

W przypadku prostej struktury możemy wykorzystać relacyjny model danych oparty o pojęcie relacji, które można uważać za pewną abstrakcję intuicyjnego pojęcia tabeli, zbudowanej z wierszy i kolumn, w której na przecięciu każdej kolumny z każdym wierszem występuje określona wartość. Baza danych jest zbiorem relacji, o następujących własnościach:

- Każda relacja w bazie danych jest jednoznacznie określona przez swoją nazwę.
- Każda kolumna w relacji ma jednoznaczną nazwę (w ramach tej relacji).
- Kolumny relacji tworzą zbiór nieuporządkowany. Kolumny nazywane bywają również *atrybutami*.
- Wszystkie wartości w danej kolumnie muszą być tego samego typu. Zbiór możliwych wartości elementów danej kolumny nazywany bywa też jej dziedziną.
- Również wiersze relacji tworzą nieuporządkowany zbiór; w szczególności, nie ma powtarzających się wierszy. Wiersze relacji nazywa się też *encjami*.
- Każde pole (przecięcie wiersza z kolumną) zawiera wartość atomową z dziedziny określonej przez kolumnę. Brakowi wartości odpowiada wartość specjalna *NULL*, zgodna z każdym typem kolumny (chyba, że została jawnie wykluczona przez definicję typu kolumny).
- Każda relacja zawiera klucz główny -- kolumnę (lub kolumny), której wartości jednoznacznie identyfikują wiersz (a więc w szczególności nie powtarzają się). Wartością klucza głównego nie może być *NULL*.

Warto dodać, że do wiązania ze sobą danych przechowywanych w różnych tabelach używa się *kluczy obcych*. Klucz obcy to kolumna lub grupa kolumn tabeli, o wartościach z tej samej dziedziny co klucz główny tabeli z nią powiązanej. Jednak w przypadku małej struktury nie musimy wykorzystywać tej możliwości, gdyż bazujemy tylko na pojedynczej tabeli.

Podstawowa struktura może wyglądać następująco, gdzie atrybutami są kolumny tabeli, natomiast wiersze tabeli (relacji) stanowią encje. Kluczem głównym relacji jest przykładowo kolumna 'ID'.

encje	atrybuty										
	ID	wzrost	zasięg ramion	tułów	długość głowy	szerokość głowy	prawe ucho	lewa stopa	lewy środkowy palec	lewe przedramię	wiek
Dawid	178.5	176	97	21.5	16.5	6	26	12	50	23	

W teoretycznym opisie modelu relacyjnego operacje na danych definiuje się w terminach tzw. *algebry relacyjnej*. Algebra relacyjna może być uważana za *proceduralny język zapytań* modelu relacyjnego. To znaczy, że dowolna informacja jaka jest do uzyskania z relacyjnej bazy danych może być wydobyta za pomocą ciągu operacji algebry relacyjnej. W praktyce w programowaniu aplikacji opartych na relacyjnych bazach danych nie korzysta się na ogół z języka proceduralnego, lecz z deklaratywnego języka opartego na tzw. *rachunku relacyjnym* (na ogół jest to SQL – Structured Query Language). Zdefiniujemy teraz schemat bazy danych w języku SQL:

- Poleceniem CREATE TABLE utworzymy podstawową relację (tabelę) o nazwie Osoby:
Składnia:
 CREATE TABLE table_name (column1 column1_type, column2 column2_type, ..., PRIMARY KEY(key_col1, key_col2, ...));
table_name to nazwa tworzonej tabeli.
column1, column2, ... to nazwy kolumn występujących w tabeli.
column1_type, column2_type, ... to domeny (zbiory wartości) poszczególnych kolumn (np. VARCHAR, INTEGER).
 Klauzula *PRIMARY KEY* określa zbiór kolumn, które stanowią klucz podstawowy tworzonej tabeli.
Polecenie:
 CREATE TABLE Osoby (Id VARCHAR, Wzrost FLOAT, Zasięg_ramion FLOAT, Tułow FLOAT, Długość_głowy FLOAT, Szerokość_głowy FLOAT, Prawe_ucho FLOAT, Lewa_stopa FLOAT, Lewy_s_palec FLOAT, Lewe_przedramię FLOAT, Wiek INTEGER, PRIMARY KEY(Id));

W każdej bazie danych wykonywane są cztery podstawowe typy operacji (CRUD – create, read, update, delete):

- dodawanie nowych danych do bazy,
- usuwanie danych z bazy,
- modyfikowanie danych zawartych w bazie,
- wyświetlanie danych zgromadzonych w

bazie. Język SQL zawiera polecenia odpowiadające tym operacjom:

- INSERT,
- DELETE,
- UPDATE,
- SELECT.

Jeśli chcemy dodać kolejne dane osób do bazy musimy wykonać następującą operację:

- Składnia:

INSERT INTO table_name SET a = value1, b = value2, ...; INSERT INTO table_name VALUES (value1, value2, ...);

Pierwsza postać polecenia pozwala określić wartości poszczególnych atrybutów poprzez ich nazwę. W drugiej postaci wykorzystywana jest domyślna kolejność atrybutów, zdefiniowana w schemacie określonej relacji.

Polecenie:

INSERT INTO Osoby SET Id = "Dawid", Wzrost = 178.5, Zasięg_ramion = 176, Tulow = 97, Dlugosc_glowy = 21.5, Szerokosc_glowy = 16.5, Prawe_ucho = 6, Lewa_stopa = 26, Lewy_s_palec = 12, Lewe_przedramie = 50, Wiek = 23; Pozostałe osoby dodajemy tak samo.

Pozostaje najważniejsza kwestia odnalezienia poszukiwanej osoby na podstawie wprowadzonego zapytania. W celu wyświetlenia danych zawartych w bazie używa się polecenia SELECT:

- Składnia (dla pojedynczej tabeli):

SELECT column1, column2, ... FROM table_name [WHERE condition] [ORDER BY o_column] [LIMIT x,y];

gdzie:

column1, column2,... to nazwy kolumn, których zawartość ma być wyświetlona jako wynik zapytania,

table_name to nazwa tabeli, z której pobierane są dane,

condition to warunek jaki muszą spełniać wiersze tabeli, aby zostać wyświetlone,

o_column to kolumna, wg. której mają być posortowane wyniki zapytania,

x to pozycja wiersza, od którego mają być wyświetlane wyniki,

y to maksymalna liczba krotek, które mają pojawić się w wyniku.

Polecenie:

Jeśli mielibyśmy taką tabelę:

Id	Wzrost	Zasięg_ramion	Tulow	Długosc_glowy	Szerokosc_glowy	Prawe_ucho	Lewa_stopa	Lewy_s_palec	Lewe_przedramie	Wiek
Dawid	178.5	176	97	21.5	16.5	6	26	12	50	23
Łukasz S.	178.5	183	92	21	17.5	6.5	26	11	49	23
Jacek	192.5	204.5	101	19.5	15.5	7	29	13	55	24
Arkadiusz	163.5	172	85	19.5	15	6	24	10	45	23
Stanisław	175.0	177.5	93.5	20.5	16	6	26	11	48	23
Przemysław	184.5	194.5	92	20	17	6	28.5	12	51	23
Damian	181.0	181.5	89.5	20	16.5	6.5	26.5	11	49	24
Piotr	184.5	195.5	94	19	17	7	29	12	52	25
Łukasz	186.5	190	95	20	16.5	6.5	27	11.5	51	25
Krzysztof W.	187.0	176	92	17.8	16	6.9	30.5	11	48	24
Michał	180.0	177	90	19	16	6.4	29	11.2	46.2	23
Piotr	191.0	192	94	20.5	16	6.5	27	12	51.3	23
Maciej	166.0	166	88	20	14	5.8	25	10.5	44	24
Łukasz K.	182.0	187.5	92	17.5	18.5	6.4	28	11.7	49	25
Krzysztof P.	168.0	186.5	80.5	18	16	5.8	26.5	11.9	50.2	24
Tomasz	186.0	178	96	22	16.5	6.2	26.2	11.3	48.2	23
Mateusz	181.3	173.5	94	21	16	6.3	27	11.2	45.7	24

Wówczas polecenie:

```
SELECT Id FROM Osoby WHERE Lewe_przedramie = 51 ORDER BY Id;
```

da w wyniku:

Id

Łukasz

Przemysław

W ten sposób otrzymaliśmy dwie osoby, dla których wartość w polu 'Lewe_przedramie' jest równa 51. Aby wynik wyszukiwania był jednoznaczny musimy sprecyzować nasze zapytanie i podać dodatkowe kryteria wyszukiwania. W przypadku tak mało licznej bazy danych jest oczywiste, że proste zapytania będą zdawały egzamin. Gdyby liczba osób znacznie wzrosła, do identyfikacji danej osoby byłoby potrzebnych więcej danych. Oczywiście jest to bardzo prosta struktura. Gdybyśmy chcieli ją choć trochę rozwinąć, trzeba by było zwrócić uwagę na dokładność dokonywanych pomiarów i uwzględnić w wyszukiwaniu granice niepewności pomiarowej (bardziej rozbudowane warunki)

4. Opis ćwiczenia - polecenia

UWAGA! Raport z badań, wyłącznie w wersji elektronicznej (rozkłady, parametry, listingi kodu Matlaba wykorzystane przy analizie oraz wnioski) proszę załączyć do sprawozdania!

1. Akwizycja danych antropometrycznych – biometrycznych

- a. *Zdjąć za pomocą przymiaru, wymiary odpowiednich części ciała zgodnie z sugestiami zaproponowanymi przez Bertillon'ego (9 pomiarów wg ilustracji 1 oraz dodatkowo wiek badanej osoby).*
- b. *W sprawozdaniu wykorzystać tabelkę z załącznika 1, zawierającą dane zebrane w roku 2008 oraz uzupełnione dane zebrane w punkcie a, która dalej będzie w prosty sposób służyła do analizy danych statystycznych.*
- c. *Określić przy każdym pomiarze zakres błędu wynikający z zastosowanego przymiaru.*

2. Badania statystyczne

- a. *Zidentyfikować próby o rozkładzie normalnym;*

Uwaga! Liczebność próby jest niewielka stąd zachodzi podejrzenie, że zidentyfikowanie prób o rozkładzie normalnym może być dosyć problematyczne. Korzystając z funkcji `normplot()` dokonaj graficznej oceny normalności rozkładów przedstawionych pomiarów. Czy jeśli wykres jest liniowy, to dane pochodzą z rozkładu normalnego? Czy w przypadku, gdy wykres traci charakter liniowy, dane pochodzą z innego rozkładu?

W Matlabie skorzystaj z funkcji `normplot()`

- b. *Estymować parametry rozkładu prawdopodobieństwa dla tych prób, które mają rozkład normalny dla zadanych poziomów ufności;*

Uwaga! Załóż trzy poziomy ufności {0,95; 0,9 oraz 0,85}. W Matlabie skorzystaj z funkcji `normfit()`, gdzie `alpha` określa poziom ufności $100(1-alpha)$*

```
[mi_e, sigma_e, mi_uf, sigma_uf] = normfit(proba, alpha);
```


Poziom ufności	Estymowana wartość średnia		Estymowane odchylenie standardowe	
0,95	Przedział ufności dla wartości średniej		Przedział ufności dla odchylenia standardowego	
0,9				
0,85				

c. Obliczyć kurtozę i współczynnik asymetrii (skośności) dla tych prób, które mają rozkład inny od normalnego;

Uwaga! Z teorii wynika, że dla rozkładu normalnego kurtoza $K=3$, dla spłaszczonego $K < 3$, dla wysmukłego $K > 3$. Gdy współczynnik asymetrii jest $A = 0$, wówczas rozkład jest symetryczny, natomiast $A < 0$ świadczy o asymetrii lewostronnej, a $A > 0$ o asymetrii prawostronnej. Czy zgadzasz się tym? Jak możesz to skomentować, przykładowy wykres?

W Matlabie skorzystaj z funkcji `kurtosis()` oraz `skewness()`

`kurtosis (proba) ;`
`skewness (proba) ;`

Próba	kurtoza	Współczynnik asymetrii

d. Obliczyć wybrane podstawowe miary tendencji centralnych tj. średnią, odchylenie standardowe i współczynnik zmienności. Istotność różnic międzyśrednimi ocenić testem t-Studenta.

Uwaga! Czy można bezpośrednio porównywać próby? Czy test istotności wymaga zrobienia dwóch pomiarów w danej próbie?

Próba	Wartość średnia	Odchylenie standardowe	Współczynnik zmienności

3. System klasyfikujący

- a. *Dla danych zebranych od kolegów z grupy zaproponować system kategoryzujący oparty o przedstawiony w części wprowadzającej i grupujący dane biometryczne zebranych osób.
(W raporcie przedstawić format rekordów bazy danych i sposób ich adresowania)*

System ma w szybki sposób selekcjonować osobę na podstawie sformułowanego zapytania. Ten punkt powinien być uzasadniony opisem teoretycznym, który należy skonstruować zbierając odpowiednie materiały z Internetu.

(W raporcie przedstawić algorytm zadawania zapytania do systemu bazy danych i algorytm przeszukiwania bazy danych)

5. Bibliografia

1. Malinowski A., Bożiłow W.: „Podstawy antropometrii”, PWN, Warszawa – Łódź, 1997.
2. Wolański N.: „Rozwój biologiczny człowieka”, PWN, Warszawa, 2005.
3. Będziński R.: „Biomechanika inżynierska. Zagadnienia wybrane”. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1997.
4. Erdmann W.S.: „Biomechanika. Przewodnik do ćwiczeń”. Wydawnictwo MAY, Gdańsk 1999.
5. Klonecki W.: Statystyka dla inżynierów. Wydawnictwo Naukowe PWN SA, Warszawa, 1999
6. Sobczyk M.: Statystyka. Wydawnictwo Naukowe PWN SA, Warszawa, 2002
7. Kordecki W.: Rachunek prawdopodobieństwa i statystyka matematyczna. Oficyna Wydawnicza GiS, Wrocław, 2003
8. Regel W.: Statystyka matematyczna w programie Matlab. Wydawnictwo „MIKOM”, Warszawa, 2003

