2. GRAFICZNA PREZENTACJA DANYCH

2.1. Histogram

Histogram

`n = histc(X, podzial)`

gdzie:

`X` – jeśli `X` jest macierzą obliczenia wykonywane są dla każdej z jej kolumn;

`podział` – wektor zawierający granice klas szeregu rozdzielczego;

`n` – wektor zawierający liczebności poszczególnych klas, klasy definiowane są w postaci przedziałów prawostronnie otwartych, ostatni element wektora zawiera ilość elementów X równych ostatniemu elementowi wektora `podzial`.

Do narysowania histogramu konieczne jest użycie funkcji: `bar(x, y, styl)`. Funkcja ta rysuje ***wykres słupkowy*** dla każdego elementu z wektora `y` w pozycji wyznaczonej przez odpowiadający mu element z wektora `x`; `styl` – w odniesieniu do histogramów wykorzystywane są style: `'hist'` i `'histc'`.

Histogram

`n = hist(X, podzial)`

gdzie:

`X`, `n` – jak przy funkcji `histc`;

`podział` – wektor zawierający środki klas szeregu rozdzielczego lub liczba określająca ilość klas – funkcja w takim przypadku automatycznie wylicza granice klas dzieląc przedział pomiędzy minimalną a maksymalną wartością z `X` na równe podprzedziały; *funkcja wykreśla histogram jeżeli jest wywoływana bez argumentu wyjściowego.*

Histogram z krzywą Gaussa

`histfit(X, podzial)`

gdzie:

`X` – jak przy funkcji `hist.` i `histc`.

`podzial` – liczba określająca ilość klas, jeżeli parametr ten zostanie pominięty ilość klas wyznaczana jest jako pierwiastek kwadratowy z liczny elementów w `X`.

Histogramy częstości i skumulowanych liczebności i częstości

Histogramy te mogą być wykreślane po przeliczeniu liczebności zwracanych przez `hist` i `histc`. Liczenie wielkości skumulowanych ułatwia funkcja: `cumsum(n)`. Funkcja ta oblicza skumulowane sumy, jeśli `n` jest macierzą sumy obliczane są dla każdej z kolumn macierzy.

Przykład 1.

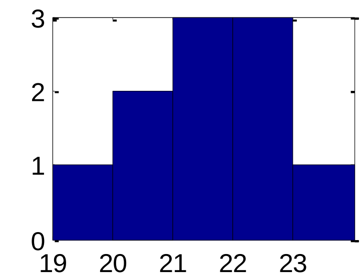
W poniższej tabeli zebrane zostały wyniki 10 niezależnych pomiarów losowo wybranego detalu (wyniki uporządkowano rosnąco). Przedstaw rozkład danych długości detalu wykorzystując histogramy i wykres normalności.

19.5	20	20.9	21.2	21.6	21.6	22	22	22.6	24

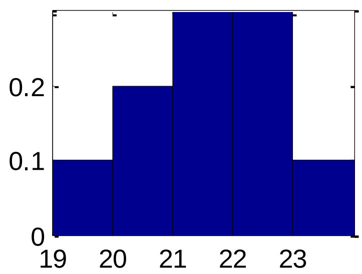
```
X=[19.5 20 20.9 21.2 21.6 21.6 22 22 22.6 24];
p = 19:1:24;
n = histc(X, p)
n =
   1   2   3   3   0   1
dl = length(n);
n(dl-1)=n(dl-1)+n(dl);

n = n(1:dl-1)
n =
   1   2   3   3   1
p = p(1:dl-1);
```

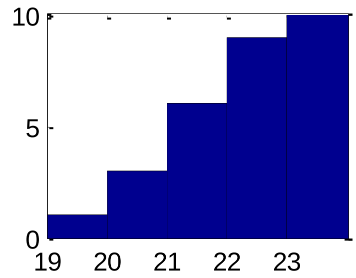
```
bar(p, n, 'histc')
```



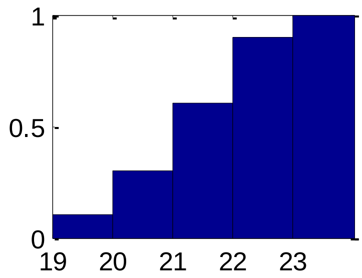

```
bar(p, n/10, 'histc')
```



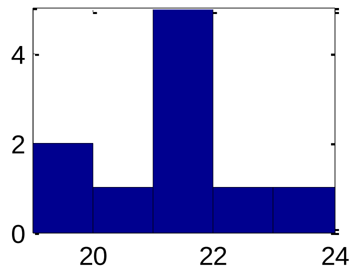
```
bar(p, cumsum(n), 'histc')
```



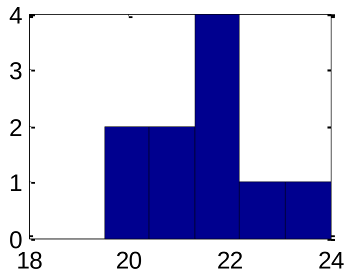

```
bar(p, cumsum(n)/10,'histc')
```



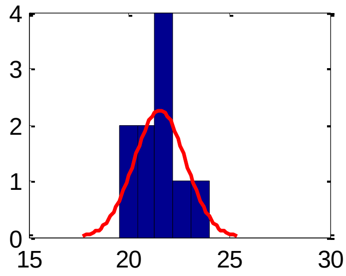

```
hist(X, 19.5:23.5)
```




```
hist(X, 5)
```




```
histfit(X, 5)
```



2.2. Wykres pudełkowy

```
boxplot(X, G, 'param', wart, ...)
```

gdzie:

`X` – wektor lub macierz, jeśli `X` jest macierzą wykres rysowany jest dla każdej kolumny macierzy;

`G` – opcjonalny wektor definiujący sposób grupowania elementów z `X`, parametr wykorzystywany jest gdy porównywane serie danych mają różną długość (nie można więc wykorzystać tych serii do skonstruowania macierzy `X`), porównywane serie danych umieszcza się wtedy jedna za drugą w wektorze `X`, wektor `G` musie się składać z takiej samej liczby elementów co wektor `X`, elementy należące do tej samej serii danych mają w wektorze `G` przypisaną tą samą wartość (np. nr serii);

`'param'`, `wart` – parametry funkcji, np.: parametr:

`'whisker'` decyduje o długościach wąsów wykresu, domyślna wartość 1.5 oznacza, że maksymalna długość wąsów wynosi 1.5 długości *rozstępu międzykwartylowego*; rzeczywista długość wąsów wyznaczana jest na podstawie elementów `X`, poszukiwane są: maksymalny i minimalny element `X` nie przekraczające zakresu wynikającego z wyznaczonej długości wąsów; wszystkie elementy przekraczające znalezione wartości zaznaczane są na wykresie jako tzw. wartości odstające.

Przykład 2.

Zmierzono średnice wałków toczonych na trzech różnych tokarkach. Otrzymano następujące wyniki:

dla pierwszej tokarki: 18.5, 18.2, 18.1, 20.3, 20.4, 20.7, 19.4, 17, 15.9, 17.1,

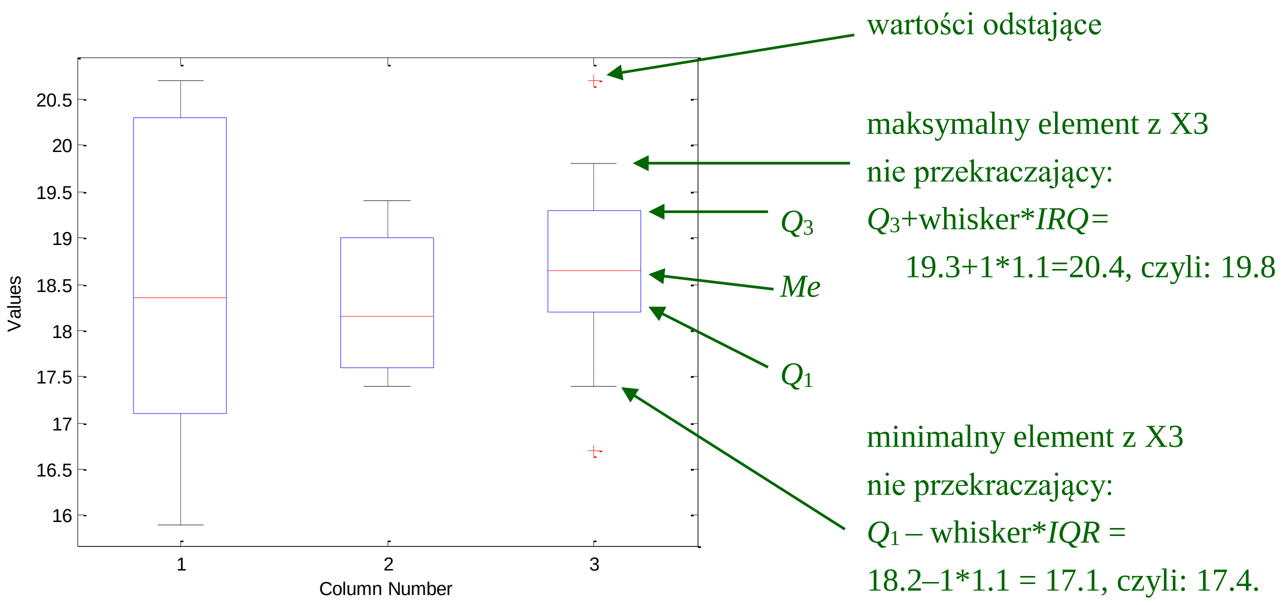
dla drugiej tokarki: 18, 17.6, 17.4, 18.5, 19, 17.6, 19.4, 18.3, 19, 17.5 i

dla trzeciej tokarki: 19, 16.7, 20.7, 18.3, 19.2, 19.8, 19.3, 17.4, 18.2, 18.2.

Porównać otrzymane wyniki na wykresie pudełkowym.

```
X1 = [18.5; 18.2; 18.1; 20.3; 20.4; 20.7; 19.4; 17; 15.9; 17.1];
X2 = [18; 17.6; 17.4; 18.5; 19; 17.6; 19.4; 18.3; 19; 17.5];
X3 = [19; 16.7; 20.7; 18.3; 19.2; 19.8; 19.3; 17.4; 18.2; 18.2];
boxplot([X1 X2 X3], 'whisker', 1)

q1 = quantile(X1, [0.25, 0.5, 0.75])
q1 = 17.1 18.35 20.3
q2 = quantile(X2, [0.25, 0.5, 0.75])
q2= 17.6 18.15 19
q3 = quantile(X3, [0.25, 0.5, 0.75])
q3 = 18.2 18.65 19.3
```



„Pudełko” wykreślone dla pierwszej tokarki jest znacznie większe od pozostałych, oznacza to, że pierwsza tokarka wykonuje detale ze znacznie większym rozrzutem.
„Pudełka” dla drugiej i trzeciej tokarki mają podobny rozmiar – druga i trzecia tokarka wykonują detale z podobnym rozrzutem. Układ obydwu pudełek odpowiada układowi z Rys. 1. b), czyli sytuacji w której można stwierdzić, że średnica wałków wykonanych na trzeciej obrabiarce jest większa niż na drugiej.

Powyższy wykres można byłoby również otrzymać wywołując funkcję `boxplot` z parametrem `G` wskazującym sposób grupowania danych:

```
X = [X1;X2;X3];
G = [1*ones(length(X1),1);2*ones(length(X2),1);3*ones(length(X2),1)];
boxplot(X, G, 'whisker', 1)
```

2.3. Wykres Q-Q

Wykres Q-Q

`qqplot(X, Y)`

gdzie:

`X` i `Y` – wektory lub macierze z danymi empirycznymi, jeśli `X` i `Y` są macierzami wykres *Q–Q* rysowany jest dla odpowiadających sobie kolumn obydwu macierzy;

Wykres normalności Q-Q

```
qqplot(X)
```

lub

```
normplot(X)
```

gdzie: `X`– jak powyżej.

Na wykresie generowanym przy pomocy funkcji `qqplot` na osi poziomej odkładane są kwantyle wynikające z teoretycznego rozkładu normalnego standaryzowanego, na wykresie generowanym przy pomocy funkcji `normplot` kwantyle teoretycznego rozkładu normalnego odkładane są na osi pionowej a oś opisywana jest nie wartościami kwantyli ale ich rzędami.

Przykład 3.

Na wykresach Q-Q porównaj:

a) rozkład normalny o parametrach wyznaczonych na podstawie danych z przykładu 1. z tym samym rozkładem teoretycznym,

b) rozkład normalny o parametrach wyznaczonych na podstawie danych z przykładu 1. z rozkładem normalnym standaryzowanym,

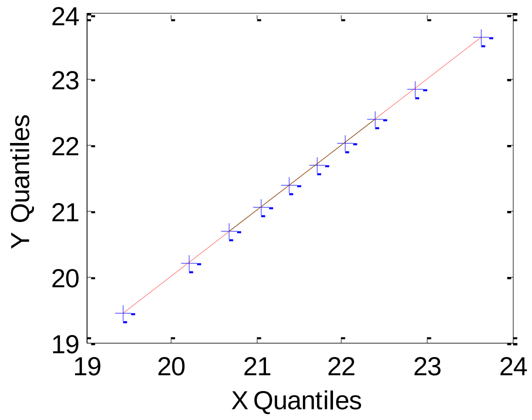
c) rozkład empiryczny długości detalu z przykładu 1. z rozkładem o parametrach wyznaczonych na podstawie danych z tego przykładu,

d) rozkład empiryczny długości detalu z przykładu 1. z rozkładem o normalnym standaryzowanym.

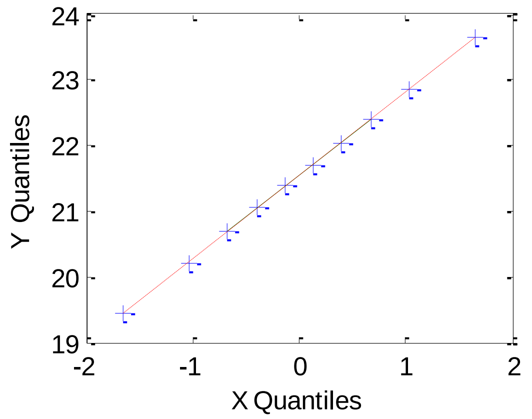
Uwaga! Rzędy kwantyli wyznaczyć wykorzystując metodę Clevelanda.

```
X=[19.5 20 20.9 21.2 21.6 21.6 22 22 22.6 24];
n = length(X)
n =
   10
i = 1:n
i =
   1   2   3   4   5   6   7   8   9   10
rzad = (i - 0.5)/n
rzad =
   0.05   0.15   0.25   0.35   0.45   0.55   0.65   0.75   0.85   0.95
Q = norminv(rzad, mean(X), std(X));
QS = norminv(rzad, 0, 1);
```

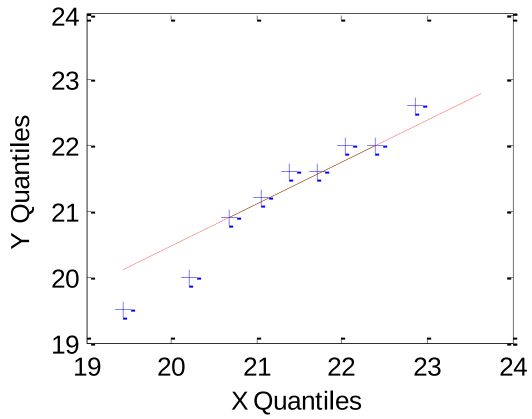
a) `qqplot(Q, Q)`



b) `qqplot(QS, Q)` *lub* `qqplot(Q)`



c) `qqplot(Q, X)`



d) `qqplot(QS, X)` *lub* `qqplot(X)`