



CO STATYSTYKA I WYKRESY MOGĄ POWIEDZIEĆ O PROCESIE?

Tomasz Demski, StatSoft Polska Sp. z o.o.

Dzięki statystyce i wykresom możemy:

- ♦ radzić sobie ze zmiennością procesów i niepewnością wiedzy,
- ♦ wydobywać użyteczną wiedzę z dużej ilości danych.

Możliwych do wykorzystania i przydatnych narzędzi jest mnóstwo, a ich praktycznych zastosowań jeszcze więcej. Zorientowanie się w dostępnych metodach, wybór odpowiedniego narzędzia i „wpasowanie” analizy danych w ramy przedsiębiorstwa ułatwia np. metodyka *Six Sigma* (por. np. [1]).

W kontekście zarządzania jakością najczęściej stosowane narzędzia to karty kontrolne, analiza zdolności procesu, badanie systemu pomiarowego (MSA) – metody te omawiamy w innych artykułach w niniejszej publikacji. Na świecie powszechnie stosuje się również planowanie doświadczeń, które w Polsce w praktycznych zastosowaniach pojawia się zdecydowanie rzadziej; bardzo ciekawe zastosowanie planowania doświadczeń omówiono w [2]. Tu zajmiemy się przykładami użycia prostych statystyk oraz wykresów.

Przykład 1

Zajmiemy się badaniem właściwości plastikowych cylindrów produkowanych na czteroskomorowej wtryskarce. Monitorowanie procesu odbywa się przez pobranie dwóch, wytworzonych w ciągu jednej godziny elementów dla każdej z czterech komór. Pomiar odbywa się co 3 godziny. Nasz przykład bazuje na analizie przedstawionej w [1].

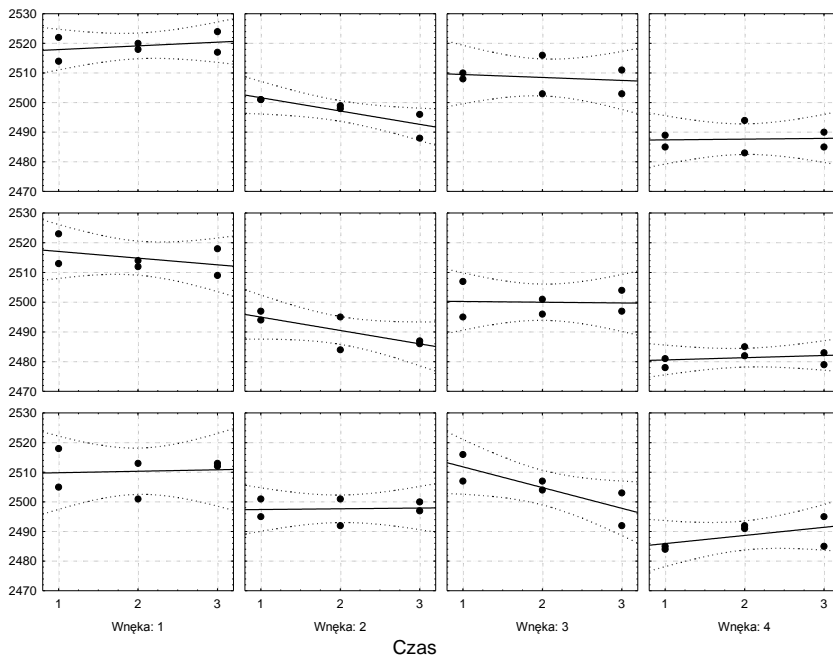
Do dyspozycji mamy stosunkowo niewiele pomiarów: są to tylko trzy serie pomiarów.

Zacznijmy od przyjrzenia się danym. Na podstawie tabeli z pomiarami trudno coś powiedzieć, ale nawet stosunkowo prosty wykres (lub zestaw wykresów) umożliwiłby uchwycenie podstawowych zależności. Na wykresie poniżej widzimy, jak rozkładają się wyniki pomiarów dla różnych części, wyprodukowanych w różnych wnękach (komorach) i w różnym czasie.

Zacznijmy od poszczególnych wykresów rozrzutu: przedstawiają one zależność od czasu średnicy produktów wyprodukowanych w ustalonej wnęce. Wydaje się, że średnice cylindrów nie zależą od czasu ich wykonania. Wystarczy „na oko” porównać rozrzut dla dwóch części wykonanych w tym samym czasie (tj. różnicę wysokości punktów o tej samej



współrzędnej X) ze zmiennością dla różnych czasów wykonania (tj. zmianę położenia pary punktów przy zmianie czasu). Jak widać, ogólnie rozrzut w obrębie pary wykonanej w tym samym czasie jest większy niż rozrzut pomiędzy tymi parami (ale dla ustalonej wnęki i miejsca pomiaru). Dla większej pewności bardziej „statystycznie” sprawdzimy jeszcze, czy występuje trend czasowy. Na wykresach średnicy względem czasu mamy dopasowaną prostą i naniesiony obszar ufności. Jeśli w obszarze tym mieści się prosta o nachyleniu zero, to możemy powiedzieć, iż dane nie wskazują w istotny sposób na występowanie trendu. Tak jest właśnie w przypadku naszych danych. Na podstawie tych obserwacji w dalszych badaniach nie będziemy traktować czasu jako istotnego czynnika (tzn. dołączymy go do składnika losowego).

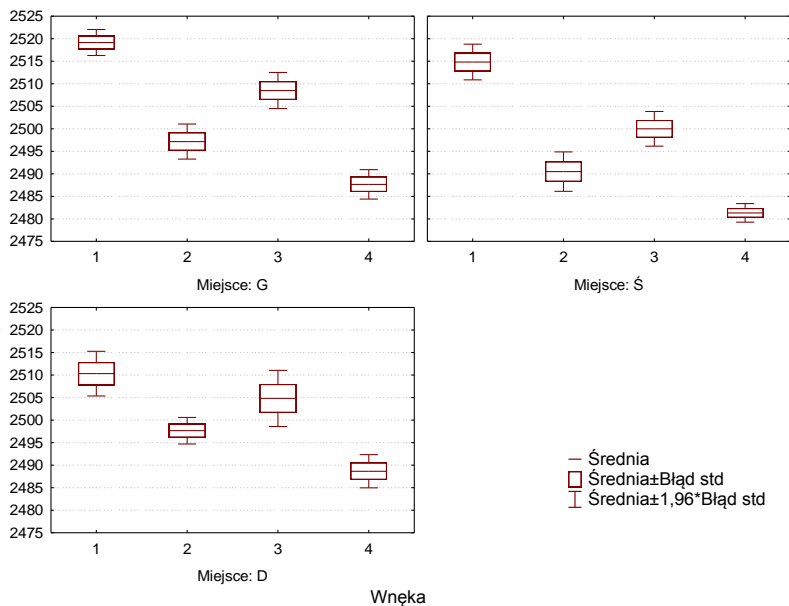


W odróżnieniu od wpływu czasu, na powyższym wykresie na pierwszy rzut oka widać wyraźną różnicę wymiarów elementów wytwarzanych w różnych wnekach. Różnice między poszczególnymi wnekami zbadamy za pomocą wykresu ramka-wąsy. Przedstawia on wartości średnie oraz przedziały $\text{średnia} \pm \text{błąd standardowy}$ i $\text{średnia} \pm 1,96 \cdot \text{błąd standardowy}$. Na rysunku poniżej dla każdego miejsca pomiaru rysowany jest osobny wykres, a oś X na poszczególnych wykresach odpowiada wnekom. Widać wyraźnie, że wneki się od siebie różnią. I tak najmniejsze średnice uzyskujemy we wnęce numer 4, a największe we wnęce numer 1. Mamy również zauważalnie odmienne rozrzuty średnic; dotyczy to zwłaszcza dolnej średnicy (*Miejsce = D*) we wnekach numer 1 i 3 – tu mamy największy rozrzut uzyskiwanej średnicy.

Najmniejszy rozrzut uzyskujemy dla wneki numer 4, zwłaszcza jeśli chodzi o średnicę mierzoną w środku wytwarzanego elementu (*Miejsce = Ś*). Przyjmując definicję jakości



zapropionowaną w klasycznym podręczniku Montgomery'ego [3]: „Jakość jest odwrotnie proporcjonalna do zmienności”, możemy powiedzieć, że najlepszą jakość mają elementy wytwarzane we wnętrze numer 4.

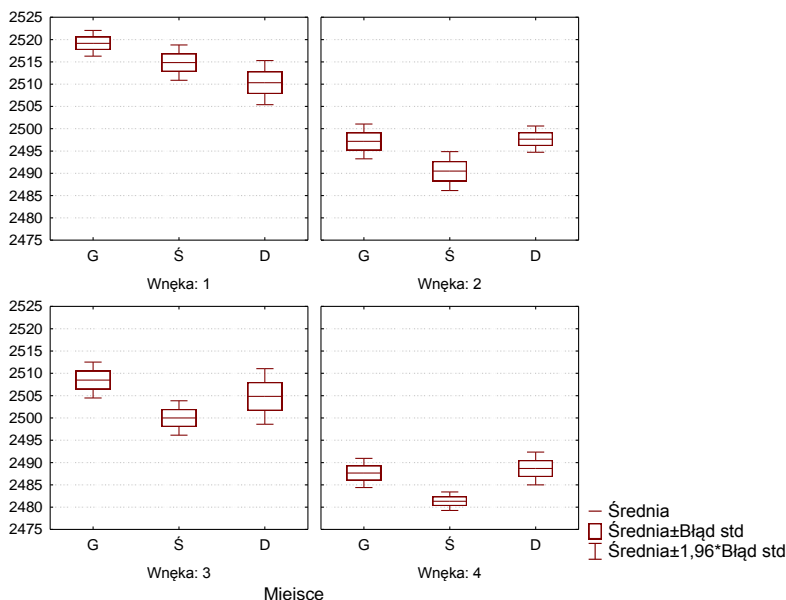


W tabeli poniżej widzimy zestawienie podstawowych statystyk dla pomiarów średnicy dokonywanych w różnych miejscach cylindra (tabelę taką możemy łatwo uzyskać za pomocą procedury *Przekroje* programu *STATISTICA*).

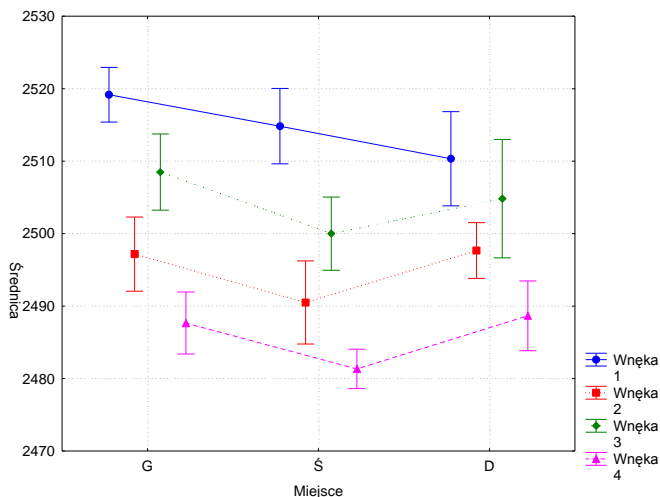
Wnęka	Miejsce	Średnia	N	Odch. std.
1	G	2519,167	6	3,60093
	Ś	2514,833	6	4,95648
	D	2510,333	6	6,18601
2	G	2497,167	6	4,87511
	Ś	2490,500	6	5,46809
	D	2497,667	6	3,66970
3	G	2508,500	6	5,00999
	Ś	2500,000	6	4,81664
	D	2504,833	6	7,78246
4	G	2487,667	6	4,08248
	Ś	2481,333	6	2,58199
	D	2488,667	6	4,58984
Ogół grup		2500,056	72	12,22776



Wykres ramka-wąsy zastosujemy teraz do sprawdzenia kształtu produkowanego cylindra – zmienimy tylko układ czynników. Na wykresie poniżej widzimy przede wszystkim, że wneki dają produkty o różnym kształcie: wneki 2, 3 i 4 dają kształt klepsydry (wąski w środku, a szeroki na brzegach), natomiast produkty wneki 1 przyjmują kształt stożkowy (szersza na górze, a węższa na dole). Niestety żadna wneka nie daje idealnego cylindra, bo średnie pomiarów w różnych miejscach różnią się więcej niż o jeden błąd standardowy.



Na wykresie poniżej widzimy średnie i 95% przedziały ufności dla średnich. Wykres ten potwierdza spostrzeżenia uzyskane za pomocą wykresu ramka-wąsy.



Wnioski uzyskane poprzez ogląd wykresów możemy (a nawet powinniśmy, por. [1]) potwierdzić za pomocą testu statystycznego. W naszym przypadku wykonamy analizę wariancji (ANOVA), aby sprawdzić, czy miejsce pomiaru i wnętrza istotnie wpływają na zróżnicowanie średnich pomiarów. Poniżej widzimy wynikową tabelę ANOVA. Bardzo silny jest wpływ *Wnęka* i *Miejsca*, ale dosyć ważna jest również tzw. interakcja tych dwóch efektów (wiersz *Wnęka*Miejsce*), tzn. poziom jednego czynnika zmienia wpływ drugiego, czyli to, co widzieliśmy na wykresach: zależność średniej średnicy od miejsca pomiaru jest odmienna dla różnych wnętrza. W przypadku interakcji prawdopodobieństwo testowe (poziom p) jest tylko minimalnie większe od 5%, czyli jeśli przyjmiemy nieco bardziej liberalny poziom istotności testu 10%, to będziemy mogli odrzucić hipotezę o tym, że interakcja nie ma znaczenia.

	Suma kwadratów	Stopnie swobody	MS	F	P
Wyraz wolny	450020000	1	450020000	18231735	0,000000
Wnęka	8301	3	2767	112	0,000000
Miejsce	504	2	252	10	0,000152
Wnęka*Miejsce	330	6	55	2	0,052495
Błąd	1481	60	25		

Przykład 2. Graficzna eksploracja danych

W drugim przykładzie zastosujemy wyróżnianie (ang. *brushing*, polska nazwa to graficzna eksploracja danych) do zbadania procesu opisywanego wieloma cechami. Wykorzystamy przykładowy plik dostarczany z programem *STATISTICA* o nazwie *IndustrialEvaporator* (por. [4]). Dane dotyczą 100 cykli produkcyjnych charakteryzowanych 8 zmiennymi.

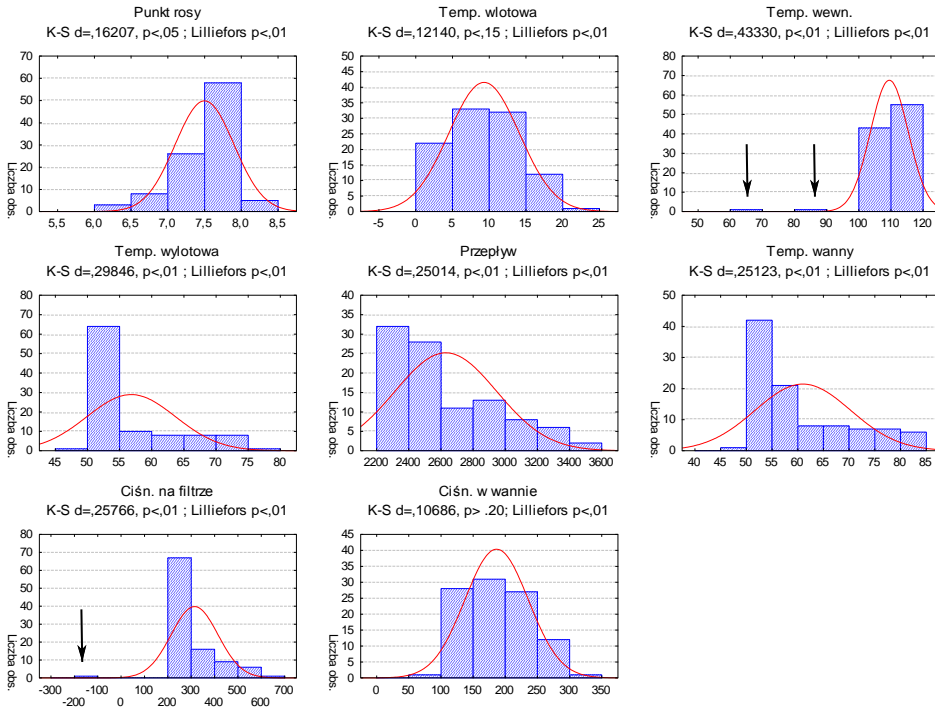
Zacniemy od wyznaczenia statystyk opisowych (są one dostępne w menu *Statystyka – Statystyki podstawowe i tabele*). Standardowy zestaw statystyk uzupełnimy jeszcze skośnością. Skośność informuje nas, czy rozkład jest symetryczny.


	N	Średnia	Minimum	Maksimum	Odch.Std.	Skośność
Punkt rosy	100	7,500	6,102	8,317	0,4000	-1,02007
Temp. wlotowa	100	9,300	1,754	20,436	4,8000	0,35068
Temp. wewn.	100	109,500	61,111	119,420	5,9000	-6,53823
Temp. wylotowa	100	56,800	49,040	75,134	6,9000	1,36184
Przepływ	100	2630,000	2336,168	3497,576	316,6000	1,07104
Temp. wanny	100	61,000	46,448	83,404	9,3000	1,10795
Ciśn. na filtrze	100	314,400	-193,888	617,610	100,2000	-0,15836
Ciśn. w wannie	100	186,500	60,089	322,941	49,4000	0,34743




W tabeli powyżej widzimy charakterystyki badanych zmiennych. Podejrzana wydaje się ujemna minimalna wartość *Ciśnienia na filtrze*. Zwróćmy uwagę, że *Temperatura wlotowa*, *Ciśnienie na filtrze* i *Ciśnienie w wannie* mają stosunkowo nieduże wartości skośności, natomiast pozostałe zmienne cechują się zdecydowanie asymetrycznym rozkładem.

Naturalnym uzupełnieniem tabeli statystyk są histogramy, które widzimy poniżej. Widać na nich bardzo wyraźnie „dziwne” obserwacje dla *Temperatury wewnętrznej* i *Ciśnienia na filtrze*. Niestety rozkład żadnej zmiennej nie jest zgodny z rozkładem normalnym – test Lillieforsa sugeruje odrzucenie takiej hipotezy dla każdej zmiennej.

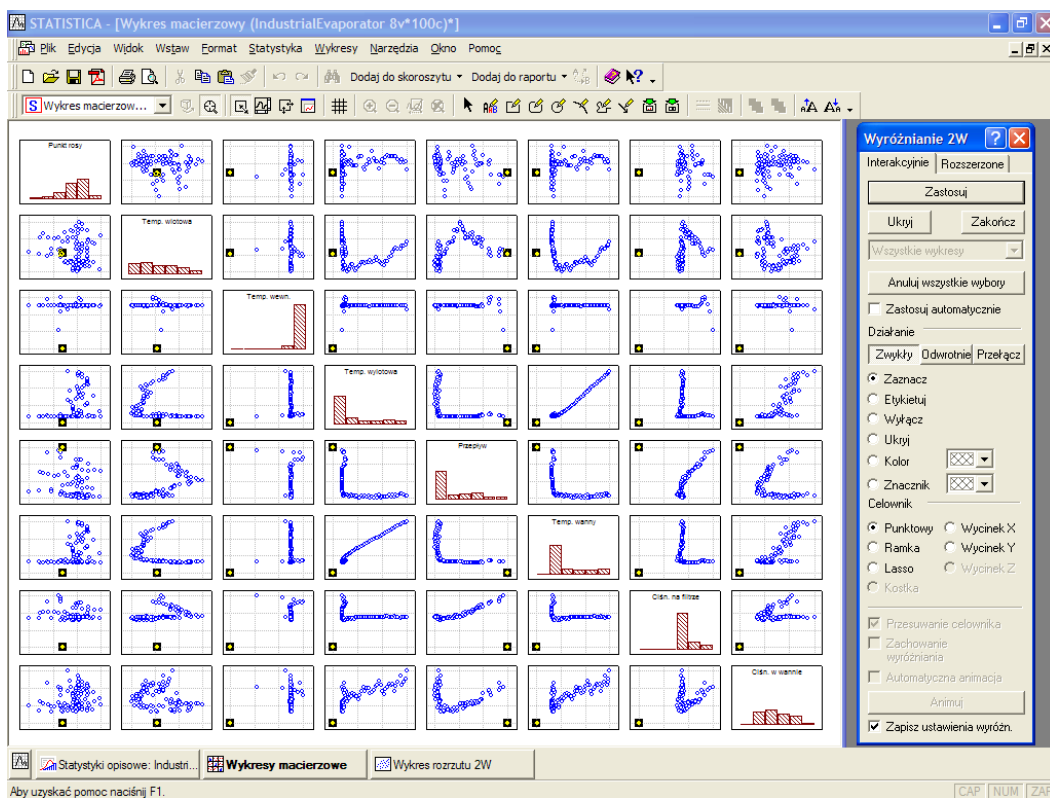


Zdecydowanie więcej dowiemy się o badanym procesie, korzystając z wykresu macierzowego, który pokazuje zależności między wszystkimi zmiennymi na jednym obrazie. Uzyskamy w ten sposób pewniejsze wnioski, bo jeśli przypadek przyjmuje nietypowe wartości dla wielu zmiennych, to jest to zdecydowanie silniejszy sygnał, że proces jest w rzeczywistości inny, niż gdybyśmy obserwowali dziwną wartość tylko w jednej zmiennej. Rysunek poniżej przedstawia środowisko *STATISTICA* z otwartym wykresem macierzowym w trybie wyróżniania (tryb ten włączamy przyciskiem  na pasku narzędzi okna wykresu).

Powyżej na wielu wykresach rozrzutu widzimy dziwne punkty odległe, a czasem bardzo odległe od „zgrupowań” typowych obiektów. Aby bliżej zbadać te punkty przechodzimy do trybu wyróżniania (naciskając przycisk ) i wskazujemy punkt, na który chcemy zwrócić uwagę. Zaczniemy od punktu, który już wcześniej uznaliśmy za „dziwny”,



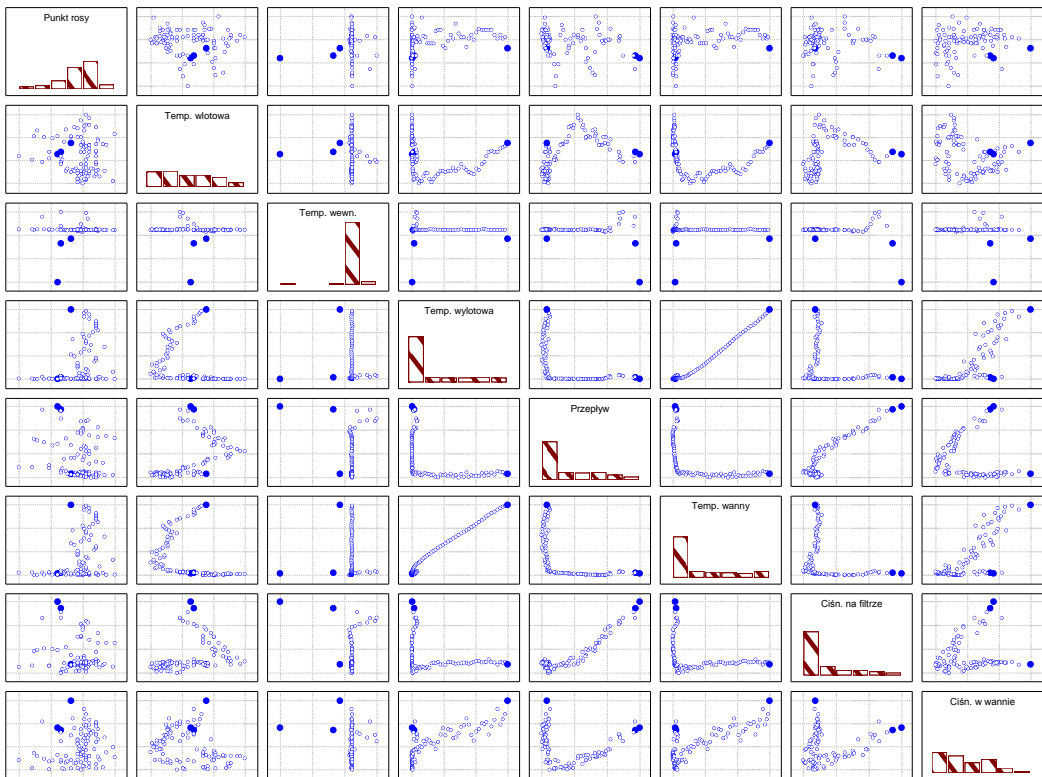
tj. obserwacji o ujemnej wartości *Ciśnienia na filtrze*. Jeśli zaznaczymy ten punkt np. na wykresie *Ciśnienia na filtrze* względem *Ciśnienia w wannie*, to program podświetli punkty odpowiadające temu procesowi na innych wykresach. Jak widać na rysunku poniżej proces o ujemnym *Ciśnieniu na filtrze* cechuje się także nietypowymi wartościami *Temperatury wewnętrznej*, *Temperatury wlotowej*, *Przepływu*, *Temperatury wanny* i *Ciśnienia w wannie*. Z dużą pewnością możemy stwierdzić, iż proces ten jest nietypowy i nie powinniśmy go brać pod uwagę przy tworzeniu podsumowań dla typowych, uregulowanych procesów. Bardzo prawdopodobne jest, że proces ten dał w efekcie produkt inny niż typowe procesy. Zauważmy, że dane dotyczą procesu chemicznego, a w takim wypadku ocena jakości wytworzonego produktu jest zwykle czasochłonna. Nasza analiza wcześniej powiadomi nas, że stało się coś niedobrego, a to pozwoli np. nie kierować nieodpowiedniego produktu do dalszej obróbki lub dostosować tę obróbkę do produktu. Zauważmy, że do zrozumienia naszego wykresu nie jest konieczna wiedza statystyczna, w gruncie rzeczy wystarczy zdrowy rozsądek.



W trybie wyróżniania dysponujemy wieloma możliwościami: np. możemy wybrane przypadki wyłączać z dalszych analiz, etykietować (aby dowiedzieć się, który przypadek wybraliśmy), ukrywać na wykresach, przypisywać mu specjalny kształt i kolor znacznika.

Dziwna obserwacja zapisana jest w 18 przypadku zbioru danych. Wyłączamy ją z dalszych analiz, aby nie „zagłuszała” innych efektów. Poniżej widzimy wykres macierzowy po

wyłączeniu podejrzanego przypadku. Widzimy kolejne trzy odstające punkty: z nietypowo niską wartością *Temperatury wewnętrznej*. Punkty te zostały zaznaczone na wykresie poniżej. Co ciekawe, punkty te odpowiadają przypadkom o numerach 17, 19 i 20, a więc sąsiadującym z procesem, który uznaliśmy za nietypowy wcześniej. Zauważmy, że co prawda na innych wykresach podejrzanego punkty nie są jaskrawo nietypowe, ale często leżą gdzieś na granicy zakresu wartości lub w miejscach, gdzie mają stosunkowo niewielu bliskich sąsiadów. Podejrzaną trójcę przyjrzymy się bliżej na wykresach rozrzutu dla wybranych par zmiennych. Wcześniej za pomocą wyróżniania ustawimy dla nich inny kształt znacznika i włączymy etykietowanie.

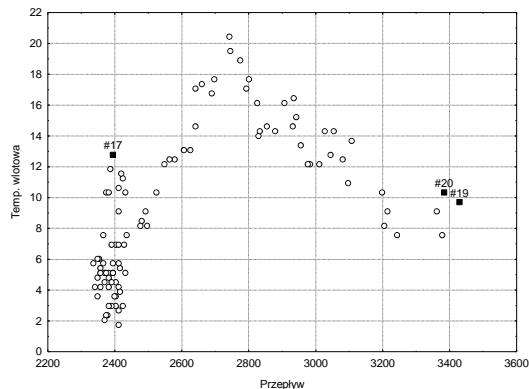
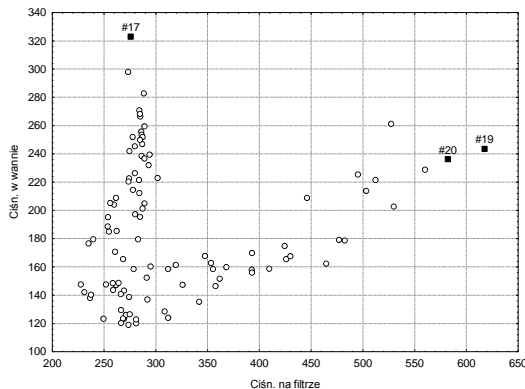


Poniżej widzimy wykresy rozrzutu *Ciśnienie w wannie* względem *Ciśnienia na filtrze* oraz *Temperatury wlotowej* względem *Przepływu*. Na pierwszym wykresie nietypowość procesów 17, 19 i 20 wynika z tego, że mają one maksymalne wartości ciśnienia. Natomiast na drugim wykresie punkt 17 ma mało sąsiadów: dla obserwowanej dla niego wartości przepływu przyjmuje on skrajną wartość *Temperatury wlotowej*. Zwróćmy uwagę, że takiej sytuacji nie byłibyśmy w stanie wykryć, biorąc pod uwagę wyłącznie rozkłady poszczególnych zmiennych.

Wydaje się, że mamy dosyć mocne podstawy do uznania również procesów nr 17, 19 i 20 za nietypowe. Po odrzuceniu tych punktów uzyskamy poniższą tabelę statystyk opisowych,



odnosząc się do typowych procesów. Średnie zmiennych nie uległy dużym zmianom, natomiast odchylenia standardowe zmieniły się dla niektórych cech dosyć wyraźnie. Przykładowo po usunięciu nietypowych punktów odchylenie standardowe *Ciśnienia na filtry* spadło ze 100,2 do 77,5. Zasadniczo zmieniła się również skośność tej zmiennej: z małej ujemnej wartości na dość dużą dodatnią. Zwróćmy uwagę, że usunęliśmy ledwie 4 przypadki na 100, a statystyki opisowe uległy wyraźnej zmianie.



	N	Średnia	Minimum	Maksimum	Odch.Std.	Skośność
Punkt rosy	96	7,515	6,102	8,317	0,4003	-1,13635
Temp. wlotowa	96	9,257	1,754	20,436	4,8848	0,36999
Temp. wewn.	96	110,393	109,075	119,420	1,6321	4,44527
Temp. wylotowa	96	56,779	52,179	74,582	6,7124	1,35561
Przepływ	96	2607,248	2336,168	3377,009	287,6483	1,02015
Temp. wanny	96	61,049	53,792	82,986	9,0435	1,12865
Ciśn. na filtry	96	314,150	227,475	559,991	77,4990	1,60082
Ciśn. w wannie	96	185,282	118,992	298,029	46,0114	0,43805

Na koniec zaznaczmy, że w rzeczywistych zastosowaniach powinniśmy nasze wnioski uzyskane w wyniku graficznej analizy danych potwierdzić, np. sprawdzając, czy procesy uznane przez nas za nietypowe dały inny produkt niż pozostałe. Odchylenia mogą być również skutkiem błędów pomiarowych i przed ostatecznym sformułowaniem wniosków to też powinniśmy sprawdzić (jednak w naszym przypadku dziwne są wartości wielu zmiennych, mierzonych różnymi urządzeniami pomiarowymi, co sugeruje, że raczej proces był nietypowy, a nie że któryś miernik uległ uszkodzeniu).

Literatura

1. Breyfogle F. W., „Implementing Six Sigma. Smarter solutions using statistical method”, John Wiley & Sons 1999.



2. Gawąd P., Ciechanowski P., „Zastosowanie statystycznych technik obliczeniowych SPC i DOE dla poprawy procesów produkcyjnych grupy ABB” (artykuł opublikowany w „Zastosowania analizy danych w przemyśle”, StatSoft Polska 2005).
3. Montgomery D. C., „Introduction to statistical quality control” John Wiley & Sons 1997.
4. StatSoft 2005, Podręcznik elektroniczny *STATISTICA* (dostępny z programem i na witrynie <http://www.statsoft.pl/download/help6.php>).