



CO TO ZNACZY WIEDZIEĆ WSZYSTKO O PROCESIE?

Adam Walanus, StatSoft Polska Sp. z o.o.

Ucz się ucz – nauka to do potęgi klucz. Albo: Zbieraj dane, zbieraj, a baza Ci sama urośnie. Co to znaczy widzieć, posiadać wiedzę? Posiadać brzmi dobrze, od przybytku głowa nie boli, twarde dyski komputerów są dziś tanie. Ale czy warto? O wartościach mówi się coraz więcej. Co warto? To jedno pytanie nie sprawia kłopotów, warto mieć zysk. Zadziwiająco stabilna koncepcja, skryształowana w momencie (dawnym) pojawienia się odpowiedniej jednostki miary.

Maksymalizacja zysku jest wielozmiennym zagadnieniem nieliniowym, trudnym do optymalizacji, choć niewyglądającym na beznadziejne, jest tu dość duży postęp. Jednak ciągle jeszcze zamiast prostych (czytaj: algorytmicznych) reguł pojawiają się mgliste pojęcia typu: system, zarządzanie, jakość, wiedza. A tak naprawdę wszystko to są liczby, prawie wszystko jest dziś cyfrowe i daje się przeliczać.

Prawie wszystko daje się mierzyć, a wynik pomiaru można zapisać. Wynik pomiaru to nie tylko liczba, to również jej znaczenie. To liczba i jej pozycja w bazie danych. Zysk, to nie tylko suma, ale i numer konta, na którym się znajduje (to żart). Czy baza danych pomiarów to wiedza? Znaczenie słowa „wiedza” – na tyle podstawowego, że trudnego do słownikowego (algorytmicznego) zdefiniowania, dość dobrze oddaje zwrot: „gdybym wiedział...”. Wiedza powinna coś umożliwiać, powinna dać się przeliczyć (oczywiście na zysk). Początki wiedzy są w liczbach, w wynikach pomiarów, ale droga do końcowych liczb (najpierw parametry procesu, potem zysk) nie jest usłana różami.

Liczby, pomiary

Czy to wszystko muszą być liczby? Dało się produkować bez liczb. W średniowieczu (choć wtedy raczej wytwarzano niż produkowano) wytwarzano produkty raczej jednostkowe, każdy inny. Następny etap, epoka przemysłowa, „stoi” na pomiarach. Człowiek zbierał wiedzę, mierząc, używając narzędzia, zyskiwał wiedzę, odczytując wskazanie miernika. Tamte liczby były nie całkiem cyfrowe. Dlatego, że to człowiek, odczytując położenie wskazówki miernika, dokonywał digitalizacji, koniecznej by wynik zapisać. Dziś, w epoce post-industrialnej, człowiek odsuwa się coraz dalej od rzeczywistości produkcyjnej. Coraz większe i wszechobecne jest narzędzie, jakim się posługuje. Jeżeli gdzieś jeszcze operator



trzyma w ręku suwmiarkę, to na pewno jest ona cyfrowa i automatycznie przesyła wynik pomiaru do bazy danych.

Miernik jest raczej częścią maszyny niż „interfejsem” między rzeczywistością a poznającym ją człowiekiem. Człowiek ma kontakt już tylko z bazą danych. Miał. Tak było jeszcze wczoraj. Dziś jest już następny etap, dziś maszyna (*data miner*) przyjrzy się bazie danych i przedstawi człowiekowi wynik analizy, pozwalający poprawić skuteczność.

Maszyna parowa jest symbolem swojej epoki, symbolem maszyny parowej bywał często regulator Watta: dwie ciężkie kulki na ruchomych ramionach, które gdy maszyna obracała się za szybko, przenosząc siłę odśrodkową, przymykały zawór, stabilizując obrót. Dziś szybkość obrotową mierzy cyfrowy miernik, a wynik po przeliczeniu przez dowolnie skomplikowaną, optymalną funkcję, steruje zaworem. Czy to jest postęp, stwierdzamy w Chabówce, gdzie jeszcze jeżdżą parowozy. Na miejsce dojeżdżamy samochodem, który nie jest już bardziej czy mniej udanym egzemplarzem, ale jest identyczny z innymi tego samego typu. Dlatego jest tak skuteczny.

Symbolem epoki po-przemysłowej mogłaby być stabilność procesów produkcyjnych, która jest kluczem do postępu (poprzez zysk).

Dlaczego procesy nie mogą być absolutnie stabilne? Dlaczego w ogóle jest sprawa stabilności? Są dwie przyczyny: (1) otoczenie i wejście procesu nie jest absolutnie stabilne, a ponadto (2) wiedza o procesie nie jest absolutna, nie jest absolutnie dokładna ani pewna. Wejście procesu jest wyjściem innego procesu. Gdyby tamten był doskonały, to część kłopotu zniknęłaby. Ale któryś proces musiałby być pierwszy, biorący prosto z natury, a ta już (jeszcze?) sterowaniu nie podlega.

R&R

Czy pomiary nie mogą być absolutnie dokładne? Nie mogą i są po temu trzy przyczyny. W kolejności, od najmniej do najbardziej podstawowej: (1) Cena miernika i pomiaru rośnie wykładniczo z poprawą dokładności, czyli mniej więcej tak, że dwa razy mniejszy błąd oznacza dziesięć razy wyższą cenę. (2) Termodynamika. Zawsze i wszędzie są szумы. (3) Mechanika kwantowa. Fundamentalna niepoznawalność, zasady nieoznaczoności itp.

Niedoskonałe pomiary mają bezpośredni wpływ na cechy produktu, poprzez układy automatyki. W ten sposób dwie główne przyczyny zmienności łączą się w jedną. Żeby przejść do następnego etapu dywagacji o wiedzy dobrze będzie założyć, co nie jest całkiem oderwane od rzeczywistości, że produkt niedoskonałego procesu może być zmierzony doskonale. A przynajmniej prawie doskonale, tak że o niedoskonałości pomiaru można zapomnieć (praktycy R&R wiedzą, o co chodzi).

Tak więc mamy w magazynie serię sztuk produktu, a w bazie danych kilka kolumn liczb – wyników pomiarów różnych właściwości owego produktu (masy, długości, twardości i szerokości). Weźmy jedną kolumnę liczb. Każda liczba w niej jest inna. Każda sztuka



niecو innej długości. Wiemy o tym i, co więcej wiemy, jaka jest długość każdej sztuki (nie tylko długość). Czy to jest cała wiedza?

Nie cała. Jaka jeszcze wiedza z kolumny kilkuset wyników pomiarów długości wynika? Wiedza statystyczna. Najwyższy etap. (Warto zauważyć, że najgłębsza przyczyna niepewności - mechanika kwantowa - posługuje się tym samym, głównym pojęciem co statystyka - pojęciem prawdopodobieństwa.)

Statystyka

Wydało się. Autor coś kombinuje, żeby przeprowadzić jakąś tezę (nawet można się domyślić jaką). Chodzi o statystykę. Paru teoretyków wymyśliło jakąś dziedzinę i usiłują nią zainteresować tych, którym jest ona zupełnie niepotrzebna. A wiadomo przecież, że w ogóle matematyka powinna być w szkołach zlikwidowania, bo i tak sobie z nią uczniowie (i nauczyciele) nie radzą. Hasło „prawda statystyczna” kojarzy się raczej z dowcipem niż z prawdziwą wiedzą. Dlatego zamiast „prawdziwą wiedzą”, lepiej zająć się wiedzą użyteczną. Pytanie, jak naprawdę wygląda nasz proces, brzmi trochę podejrzenie, ale wiedzieć wszystko o procesie chciałby każdy dobrze umotywowany inżynier produkcji. Oczywiście „wszystko” oznacza „wszystko, co można wiedzieć”, zakładając, że pensja jasnowidza przekracza ekonomiczne możliwości firmy.

Czy jakaś wiedza ukryta jest między wierszami? Zawsze tak jest, a bynajmniej nie trzeba wróżki, by ją ujawnić. Jeżeli pewna sztuka produktu ma długość 12,34mm, to ma długość 12,34mm i koniec. Jeżeli inna sztuka produktu ma 12,39mm, to ma 12,39mm, ale teraz są dwa wiersze tekstu, a więc jest też przestrzeń między wierszami, a w niej informacja, że dwie sztuki produktu różnią się o 0,05mm. A to już jest COŚ! Proces daje produkt, którego długość zmienia się o 0,05. Co to dokładniej znaczy? Czy kolejne sztuki będą miały raz 12,34, a raz 12,39? Może tak, ale raczej nie. Na razie wiadomo, że różnica może wynosić 0,05. Jeżeli obie zbadane sztuki produktu nie były identyczne i wykryliśmy różnicę, to zapewne ta różnica też będzie mogła być różna. Jak już jest bałagan, to do końca (to się nazywa „normalność”).

o o o o O o o O O O o o o o o o o o o o

Warto przy okazji zwrócić uwagę, że proste i niebudzące wątpliwości zdanie, że dwie sztuki produktu mogłyby być identyczne, jest z gruntu fałszywe. Jeżeli już była tu mowa o mechanice kwantowej, to można do rzeczy podejść ściśle: otóż identyczne są dwa elektrony, dwa atomy, dwie cząsteczki H₂O, itd., ale nie dwa obiekty składające się z miliarda miliardów atomów. Każde dwa (tzw. makroskopowe) przedmioty są różne.

Zdolność

Z reguły klient zamawia coś konkretnego. Oczywiście bardzo dobrym interesem może być sprzedaż produktów *hand made*, gdzie klient przyglądał się będzie dwóm filiżankom



i z satysfakcją stwierdzi, że są różne. Nie o ten segment rynku jednak w tym przypadku chodzi. Raczej o produkcję przysłowiowej tulejki, która ma gdzieś wejść (bez luzu). Klient chciałby średnicę 12,35mm. Co to matematycznie oznacza? Matematycznie liczbę 12,35 można też zapisać jako 12,3500000.... Wyprodukowanie tulejki o specyfikacji $\Phi=12,3500000\text{mm}$ jest fizycznie niemożliwe, bo nie da się przyciąć na pół atomu Fe. Praktycznie również wykonanie tulejki $\Phi=12,35\text{mm}$ może być trudne albo ekonomicznie nieuzasadnione. Dlatego dochodzi się z klientem do kompromisowego porozumienia. Klient zgadza się, że nie każda sztuka musi mieć dokładnie 12,35mm. Ale przecież nie może mieć 12,50mm. Jak wyrazić dopuszczalną zmienność?

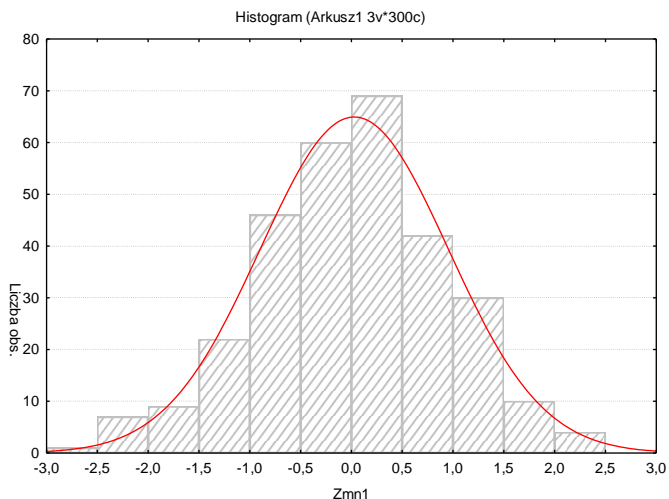
Odpowiedzialny inżynier u klienta powie, że najlepiej żeby tulejka miała 12,35, a jak już musi być mniejsza, to żeby nie przekroczyła 12,32, oraz żeby w żadnym wypadku nie była grubsza od 12,36. Jednak gdy dział handlowy dowie się, ile za taką tulejkę trzeba zapłacić, to wymusi na swoim koledze z produkcji większą tolerancję. I tak powstaje kompromis. A kompromis najlepszy jest, gdy strony się porozumieją. Oznacza to, że klient powinien uwzględnić technologię produkcji zamawianego wyrobu.

Technologia tulejki (w aspekcie jej średnicy) nie jest wyjątkowa, zdecydowana większość liczbowych cech najróżniejszych produktów charakteryzuje się taką samą, pewną podstawową właściwością. Co to za właściwość? Rzecz nie jest banalna.

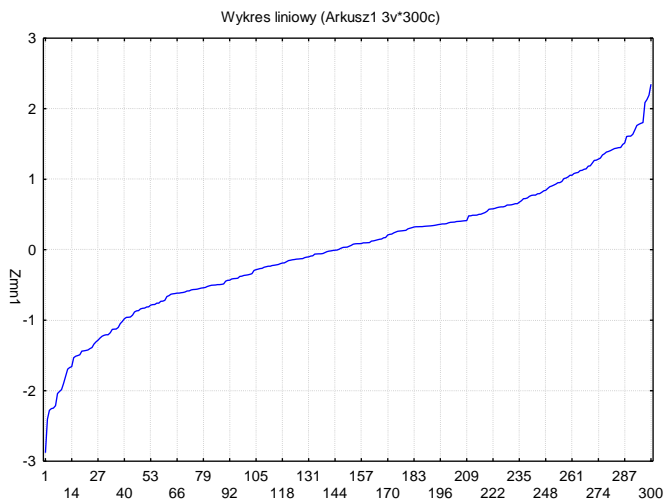
Normalność

Co wspólnego mają ze sobą: średnica tulejki, jej długość (56,7mm), masa tulejki, masa sworznia, głębokość otworu w sworzniu, masa opony, jej twardość, zawartość siarki w gumie, wilgotność piasku, pH napoju chłodzącego i jego objętość w butelce? Cechy te mają najróżniejsze wartości liczbowe i najróżniejsze wymiary (mm, kg, mg, pH, Nmm-2, HV, itd.). Wspólna właściwość większości tego typu cech produktów, o której tu mowa, jest właściwością statystyczną. By ją zobaczyć (a potem zrozumieć), musimy dysponować serią pomiarów danej cechy, wykonanych dla różnych sztuk produktu. Im więcej pomiarów tym lepiej; w poniższym przykładzie jest 300 pomiarów. Postępujemy w przedstawiony poniżej sposób.

(1) Algorytm szybszy. W *STATISTICA* klikamy *Wykresy/Histogramy*, wybieramy zmienną i klikamy OK. Otrzymujemy wynik jak na poniższym rysunku, którego omówienie zostawmy na później.



(2) Algorytm ręczny, podstawowy. Porządkujemy serię wyników w kolejności rosnącej i tworzymy zwykły wykres liniowy.



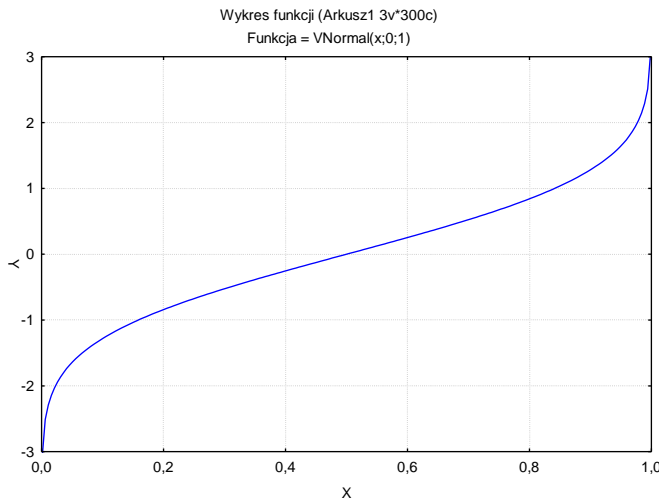
Algorytm drugi ma tę zaletę, że wszystkie kroki są w nim widoczne jak na dłoni i są zrozumiałe. W algorytmie pierwszym program coś robi, coś sam dobiera itd., ale końcowy wynik jest czytelniejszy. Wszyscy wiedzą, o co chodzi na pierwszym rysunku.

Tytułowa wiedza o procesie, jak każda wiedza, zakłada rozumienie. Nie ma wiedzy bez częściowego przynajmniej rozumienia jej składników i relacji między nimi zachodzących. Rozumienie zawsze jest ograniczone, bo zawsze można rozumowanie doprowadzić do pytania: dlaczego TO wszystko się zaczęło (15 mld lat temu). Ważne jest wiedzieć, co rozumiemy, a czego już nie możemy rozumieć, a więc i nie musimy (z technologicznego punktu widzenia).



Na drugim rysunku, którego konstrukcja jest jasna i prosta, widać podstawową cechę analizowanych danych, jednak jej uchwycenie nie jest już tak jasne i proste. Chodzi o kształt otrzymanej krzywej. Otóż ten kształt, w jego podstawowych zarysach, będzie taki sam dla twardości gumy, długości trzpienia i zawartości cukru w soku (inne będą tylko liczby na osi pionowej).

Ideał (w pierwotnym sensie tego słowa – czyli esencja uwolniona od cech nieistotnych) wygląda tak:



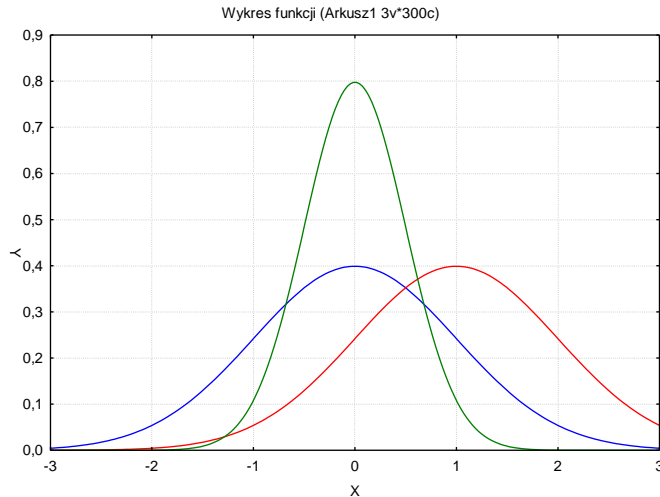
Jeżeli na powyższym wykresie zamienić osie miejscami i obliczyć pochodną funkcji, to otrzyma się znaną (i lubianą) krzywą dzwonową (patrz pierwszy rysunek), czyli krzywą rozkładu normalnego. Wyniki pomiarów mają rozkład normalny. Kropka.

To jest wiedza. Wiedzieć, że nasz proces podlega rozkładowi normalnemu (w aspekcie pewnej cechy produktu), to wiedza użyteczna. Dwa punkty uzupełnienia. Wiedza ta jest wiedzą bo: (1) jednak nie wszystkie procesy opisać się da rozkładem normalnym i musimy wiedzieć, czy nasz akurat tak, oraz (2) jeżeli mamy rozkład normalny, to otwarta jest droga do uszczegółowienia wiedzy, gdyż rozkład normalny ma dwa parametry, które, po prostu trzeba, dla konkretnej cechy produktu, w konkretnej chwili, znać.

Skomplikowany jednak kształt krzywej dzwonowej (ludwisarze wiedzą, że nie jest on banalny) mieści się w matematycznej formule $y=\exp(-x^2)$. Dlaczego tak jest, dlaczego taka właśnie funkcja opisuje procesy produkcyjne (i nie tylko!), to zaczynają się sprawy trudniejsze, tu można się zastanawiać, co to znaczy rozumieć. Początek jest względnie prosty, zakłada się, że jest bardzo wiele, raczej drobnych, a niezależnych, różnych przyczyn, że długość trzpienia odbiega nieco od 12,35mm. Tylko z tego założenia, na podstawie ścisłego, matematycznego dowodu centralnego twierdzenia granicznego rachunku prawdopodobieństwa wynika przytoczony wzór. Jeżeli któryś inżynier produkcji „przegryzłby się” przez ten dowód, to mógłby powiedzieć, że rozumie rozkład normalny. Formalnie mógłby tak powiedzieć (mógłby też wystąpić o podwyżkę pensji), ale, czy



byłoby to rozumienie w potocznym sensie: „aaa, teraz rozumiem”? Wątpliwe. Raczej nie zaleca się studiowania, dla celów SPC, podstaw statystyki matematycznej. Jednak wykreślić sobie raz w *STATISTICA* kilka przykładów funkcji $y=1/(\sqrt{2*\pi}*\sigma)*\exp(-0,5*(x-mi)^2/\sigma^2)$ nie byłoby złe.



Co jest obowiązkowe, by móc wiedzieć wszystko o procesie. Pamięciowe opanowanie powyższej formuły niekoniecznie, ale dwa parametry w niej występujące - tak. Parametr *mi* to średnia procesu, a *sigma* to sigma procesu (uprasza się statystyków matematycznych o zachowanie dla siebie uwag terminologicznych). Każdy proces, w każdym momencie ma jakąś średnią i jakąś sigmę (procesy niegaussowskie mają jeszcze skośność i kurtozę). Oba te parametry niestety ciągle się wahają i ich dokładnych wartości nie znamy i znać nie będziemy. Wiedzieć coś jednak o nich musimy.

Karta kontrolna (pomiarowa, przy ocenie liczbowej) daje wiedzę o średniej procesu i jego sigmie. Co więcej, daje wiedzę o stabilności średniej i sigmy. Karta kontrolna *STATISTICA* powie nam też, czy rozkład jest normalny, ale będzie to wiedza, a nie informacja. Podobnie zresztą jest ze stabilnością (uregulowaniem) procesu. Nie otrzymujemy prostej informacji TAK/NIE, uregulowany/nieuregulowany, normalny/niegaussowski. Na tym etapie rzeczy się komplikują, i choć jednobitową (0/1, dobry/niedobry) informację wygenerować można, to jednak program nie ludzi użytkownika, dostarczając mu mniemaną wiedzę.

Wartości parametrów, średniej i sigmy mamy na karcie kontrolnej po prostu podane liczbowo. Jeżeli na karcie nie ma żadnych czerwonych znaków (sygnałów o prawdopodobnym rozregulowaniu), to podane wartości parametrów są pełnowartościowe, w przeciwnym wypadku do pełni wiedzy o procesie potrzeba też wiedzy o jego rozregulowaniu. Korzystając z najprostszej automatyki kart kontrolnych *STATISTICA*, dostaniemy prostą informację typu TAK/NIE, w ten sposób na przykład, że karta kontrolna pojawi się na ekranie monitora dopiero w momencie wystąpienia rozregulowania. Wtedy oczywiście zaczynają się „schody”. Pozostając na razie na parterze, zauważyć trzeba, że brak



rozregulowania pozwala na wykorzystanie średniej i sigmy procesu do nabycia o nim pełnej wiedzy.

Pełna wiedza o procesie też może się sprowadzać do jednobitowej informacji typu TAK/NIE, konkretnie ZDOLNY/NIEZDOLNY. Tym bardziej, jeśli szczęśliwym trafem akurat jest zdolny (do spełnienia wymagań klienta). Wiedzy nie potrzebujemy, gdy wszystko jest w porządku, lecz jest niezbędna, gdy trzeba naprawić, a nie wiadomo co (wiadomo, że proces się popsuł, ale co w nim?). Jak zdolność procesu oblicza się ze średniej procesu, sigmy, oraz klienckiej specyfikacji, wiadomo, jest na to wzór (dwa wzory, no może cztery; Cp, Cpk, Pp i Ppk). W zasadzie jeżeli zdolność wychodzi większa od jeden, to proces jest zdolny, w przeciwnym wypadku nie, i mamy jeden bit TAK/NIE. Proces, który tak działa, że zapominamy o nim, dopóki (raz na parę miesięcy) na ekranie pojawi się wielkie NIE, jest doskonały. W praktyce jednak bezpieczniej jest wiedzieć, jaka (liczbowo) jest zdolność (ta i inna).

Wiedza o procesie

Wiedza o procesie to:

- ◆ znajomość średniej procesu dziś, wczoraj i w perspektywie historycznej,
- ◆ znajomość sigmy procesu z bieżącej karty kontrolnej i z przeszłości,
- ◆ znajomość pojęcia normalności rozkładu, wynikającego z odnotowywania skośności i kurtozy (zakładka *Niegaus*. karty kontrolnej pojedynczych obserwacji), a także obserwacji histogramu,
- ◆ usystematyzowana znajomość charakteru występujących rozregulowań,
- ◆ znajomość wymagań klienta.

Klient na końcu. *Last but not least*. Nie ma sensu pytać o zdolność procesu, dopóki nie jest on uregulowany.

MSPC

Wszystko to dotyczy jednej cechy produktu. Tak zwanej kluczowej lub krytycznej cechy. Ale produkt ma wiele cech, może mieć nawet wiele cech kluczowych. Również jeżeli produkt ma rzeczywiście tylko jedną cechę, to wiele cech ma proces. Bardzo ważna dla wiedzy o procesie może być znajomość temperatury pieca, a nie tylko końcowa wytrzymałość wygrzewanego produktu.

Wspominana już współczesna tendencja do zapełniania dysków danymi powoduje, że SPC musi poradzić sobie z wielością cech produktu/procesu. I radzi sobie, jest *STATISTICA* MSPC – Multi SPC, czyli *Wielowymiarowe SPC*.



The screenshot shows the STATISTICA - MSPC.sta software interface. The main window displays a data table with columns labeled 1 through 11. The first three columns are 'Data', 'Czas', and 'Kod'. The remaining columns are 'Masa', 'Twardość', 'Wytrzymałość', 'Wilgotność', 'Wymiar 1', and 'Wymiar 2'. The data rows show values for these variables across different dates and times. A menu is open over the table, listing various statistical analysis options such as 'Statystyki podstawowe i tabele', 'Regresja wieloraka', 'ANOVA', 'Statystyki nieparametryczne', 'Dopasowanie rozkładów', 'Zaawansowane modele liniowe i nieliniowe', 'Wielowymiarowe techniki eksploracyjne', 'Statystyki przemysłowe', 'Analiza mocy testu', 'Sieci neuronowe', 'Data-Mining', and 'QC data mining i analiza przyczyn'. The 'Statystyki przemysłowe' option is highlighted.

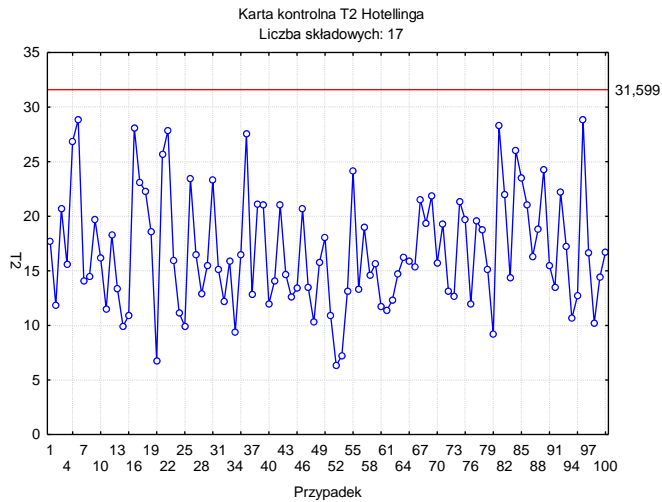
	1	2	3	4	5	6	7	8	9	10	11
	Data	Czas	Kod			Masa	Twardość	Wytrzymałość	Wilgotność	Wymiar 1	Wymiar 2
2005/10/01	2005/10/01	10:05	0,057	46	0,822	0,446	0,487	0,958	0,818	0	0
2005/10/01	2005/10/01	10:07	0,497	14	0,794	0,152	0,742	0,505	0,633	0	0
2005/10/01	2005/10/01	10:11	0,068	52	0,470	0,671	0,896	0,091	0,413	0	0
2005/10/01	2005/10/01	10:11	0,314	60				0,547	0,242	0	0
2005/10/01	2005/10/01	10:11	0,803	61				0,478	0,666	0	0
2005/10/01	2005/10/01	10:11	0,579	62				0,278	0,512	0	0
2005/10/01	2005/10/01	10:11	0,171	63				0,389	0,504	0	0
1900/01/00	1900/01/00	10:11	0,689	64				0,908	0,446	0	0
1900/01/00	1900/01/00	10:11	0,747	65				0,267	0,100	0	0
1900/01/00	1900/01/00	10:11	0,429	66				0,447	0,081	0	0

Jeżeli są dane, bo są mierniki i wszystko mierzą, to nie ma innej możliwości, jak wykorzystać tę informację (bo zrobi to u siebie konkurencja).

Jedna cecha produktu ma rozkład normalny (albo inny), o danych parametrach. Dwie cechy, poza tym że każda ma swój rozkład, mają wzajemną korelację. Dwie cechy mogą być (przeważnie są) od siebie uzależnione. Cecha A może zależeć od B, albo odwrotnie, lub obie zleżeć mogą od cechy C (cechą C procesu może być np. temperatura powietrza albo po prostu czas). Korelacje są dodatnie (A większe, to i B większe) i ujemne (np. czas suszenia i wilgotność). Oczywiście korelacje mogą być silne i słabe, warto przypomnieć, że współczynnik korelacji może maksymalnie wynosić 1 (albo -1, przy ujemnej korelacji) albo być bliski zeru, co oznacza brak zależności.

Jeżeli mamy N cech, to mamy $N*(N-1)/2$ korelacji (każda z każdą, ale tylko raz i nie sama z sobą). Przy $N=30$ mamy 435 korelacji; jak nad tym zapanować, jak 435 współczynników korelacji przerobić na wiedzę? W *STATISTICA MSPC* służy do tego, na przykład *STATISTICA PCA*. PCA (*Principal Components Analysis*), czyli analiza składowych głównych, jest znaną od dość dawna metodą matematyczną, wykorzystywaną szeroko w statystyce. W jej nazwie tkwi założenie, że są jakieś główne składowe całej zmienności procesu, manifestującej się w naszych N zmiennych. Jeżeli zmienne są skorelowane, to tak w istocie jest, wtedy PCA robi z tych korelacji użytek. Jeżeli korelacji nie ma, to jest niedobrze. PCA nic nie da, bo proces jest rzeczywiście bardzo skomplikowany, i każda z jego N cech pokazuje coś innego, albo pomiary nie spełniają zasad R&R i wprowadzają szum maskujący prawdę o procesie (mierniki z zasady nie są ze sobą skorelowane). Najczęściej jednak PCA wyodrębnia kilka najważniejszych składowych, takich że zawierają one większość informacji, jaką niosą wszystkie mierzone zmienne.

Jak działa PCA, jak obraca się macierze kowariancji, tego nie musimy rozumieć. Coraz więcej operacji na danych będzie wykonywanych poza naszą świadomością. Autor nie podejmuje się rozstrzygać, na ile trzeba rozumieć, co to jest składowa główna (jaki jest jej sens i co to znaczy, że jest ona kombinacją liniową zmiennych). Karta kontrolna Hotellinga, typowy produkt MSPC, wygląda na łatwą w użyciu (wykres niżej), ale jej nazwa brzmi podejrzanie i podejrzenie to jest słuszne. Im coś jest łatwiejsze w użyciu, tym bardziej jest skomplikowane w środku.



Na pewno lepiej jest rozumieć więcej, niż mniej, jednak nie da się obliczyć, ile warto wydać na szkolenia statystyczne. Ocenieć to można co najwyżej po wstępnych doświadczeniach, przy czym z faktu, że wstępne uruchomienie MSPC jest zachęcające, wyciągnąć można wniosek, że wystarczy już szkoleń, jak i że należy je kontynuować (bo rzecz jest dobra).

Maszyny się uczą

Postęp technologii w epoce informacji powoduje, że jest coraz więcej danych, coraz więcej liczb, czy innych kodów. Na szczęście równolegle rozwijane są techniki analizy wielkich zbiorów danych. Człowiek nie nadąza za rozwojem komputerów, nigdy nie będzie tak szybko liczył, ale z drugiej strony komputer nigdy nie będzie miał wszystkich informacji, jakie ma zespół inżynierów. Co więcej, jest wiele takich działów przetwarzania danych (np. rozpoznawanie obrazów), gdzie komputery w ślimaczym tempie próbują dogonić możliwości mózgu ludzkiego. Maszyna (uczenie maszyn, *data mining*, MSPC) może pomóc, jest wręcz konieczna (wg zasady, że wszystko co jest do dyspozycji musimy zastosować, w ramach rachunku ekonomicznego, bo zrobią to inni), ale pozostaje jeszcze miejsce dla zwykłej ludzkiej wiedzy.

By wiedzieć wszystko o procesie, trzeba mieć odpowiednie narzędzia statystyczne i umieć wykorzystać dostarczaną przez nie informację. Jedno i drugie wiąże się z kosztami, ale wygląda na to, że postęp ciągle idzie w jednym kierunku (który niełatwo nazwać) i stanie w miejscu spowoduje, że coraz trudniej będzie nadążyć za liderami. Przy tym jest tu szansa dla nieco zapóźnionych, bo koszt oprogramowania statystycznego, mimo wszystko nie jest porównywalny z kosztem maszyn czy lepszych surowców, a coraz więcej zależy dziś od informacji i wiedzy.