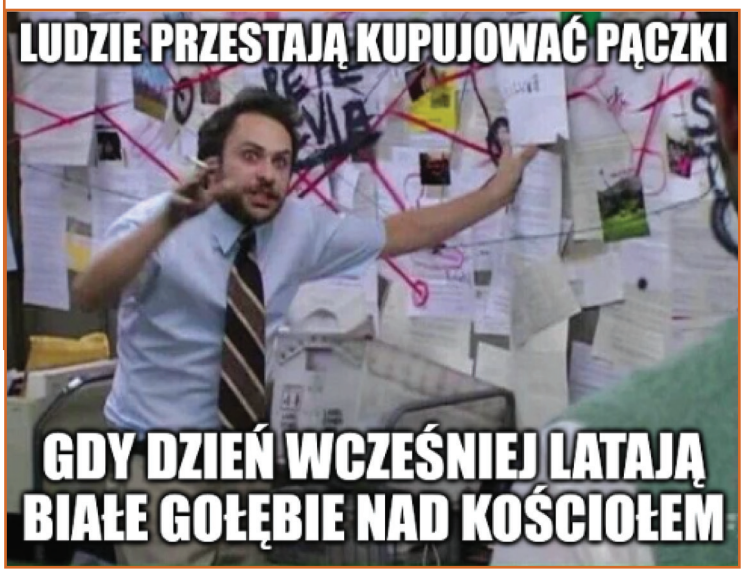
Zastosowanie prostych testów statystycznych w piekarni



Czas guseł i wierzeń w nadprzyrodzone moce minął bezpowrotnie, wydawałoby się na zawsze. Ale czy na pewno?

Nieraz spotykałem doskonale wykształconych inżynierów, którzy wierzyli, że „dwa dni po pełni księżyca zużycie gazu zawsze jest większe o 12%", albo opowiadali, że „pomalowanie maszyny na czerwono przynosi pacha". Każde takie spostrzeżenie jest cenne, jednakże, aby w nie uwierzyć należy najpierw zweryfikować je za pomocą testu statystycznego.

Sprawdzenie testem „opowieści z mchu i paproci" jest bardzo proste. Wystarczy zebrać dane i wpisać je do darmowego kalkulatora w internecie.

Jeżeli coś sprawdzimy statystycznie, możemy dalej o tym rozmawiać, tym razem już poważnie i nikt nie będzie nas brał za bajarzy opowiadających o „pełniach księżyca" i „tajemniczych mocach". Pomijam już fakt, że powołanie się na wyniki testów statystycznych może spotkać się z uznaniem ze strony analityków czy księgowych.

Piekarnia to zakład, a nie szkoła czarnoksiężników

Jednymi z najważniejszych umiejętności w zarządzaniu piekarnią jest wnioskowanie i przewidywanie. Na podstawie własnego doświadczenia wiemy na przykład, że chleb najlepiej sprzedaje się w poniedziałki, środy i soboty. Dzięki tej wiedzy planujemy taki poziom produkcji, który spowoduje minimalny poziom zwrotów. Niestety nie zawsze udaje nam się trafić z produkcją w punkt. Każdy zbyt duży poziom produkcji to tysiące złotych wyrzucone w błoto. No może nie do końca w błoto, ale w bułkę tartą, surowiec do zagęszczenia ciasta lub w paszę dla drobiu. Jeżeli wyprodukujemy za mało pieczywa narazimy się lokalnej społeczności. Każdy piekarz, który miał taki epizod wie co to znaczy narazić się połowie miasteczka.

A więc warto się przyłożyć do przewidywania przyszłości. Wiemy na przykład, że pieczywo sprzedaje się lepiej w określone dni tygodnia. Żeby lepiej dowiedzieć się ile należy produkować, można wyciągnąć średnią wielkość sprzedaży ze wszystkich poniedziałków. Otrzymana liczba będzie na tyle dobra na ile dobre będzie towarzyszące tej liczbie odchylenie standardowe. Jeżeli dowiemy się, że średnio w poniedziałki sprzedajemy 1000 bochenków chleba z odchyleniem plus minus 250 bochenków, to co mamy z tą informacją zrobić? Czy mamy wyprodukować 1250 bochenków, a jak sprzedaż wahnie się w dół to zostaniemy ze zwrotami na poziomie 500 bochenków. Czyli zwroty będą na poziomie połowy średniej sprzedaży ze wszystkich poniedziałków tego roku.

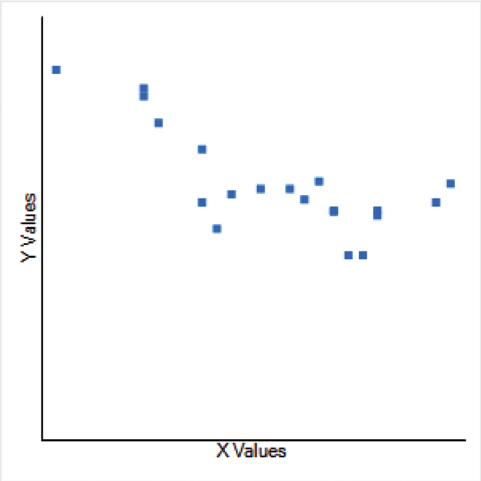
Nie tylko dzień tygodnia

Z powyższego przykładu wynika, że na sprzedaż chleba wpływają jeszcze inne czynniki. Po namyśle dochodzimy do wniosku, że wpływ ma też pogoda. Konkretnie poziom odczuwalnej temperatury, wilgotność powietrza, poziom opadów. Pojawia się szereg kolejnych czynników. Na przykład zbliżający się długi weekend lub początek okresu wakacyjnego. Wyraźnie też widać zmiany popytu pomiędzy okresem przed dziesiątym dniem każdego miesiąca i po dziesiątym dniu. Ta enigmatyczna zależność wynika po prostu z terminu wypłat wynagrodzeń w kilku okolicznych fabrykach.

Co możemy zrobić po zebraniu tych informacji. Jeżeli ktoś jest cierpliwy może poszukać korelacji pomiędzy poziomem temperatury a poziomem sprzedaży. Pomiar korelacja jest możliwy gdy mierzymy dwie zmienne ciągłe. Zmienna ciągła to zmienna, którą można zapisać używając ułamków lub gdy zmienna występuje w dużych ilościach. Załóżmy, że chcemy zmierzyć poziom korelacji pomiędzy sprzedażą chleba a temperaturą. Podstawiamy dane do darmowego kalkulatora w internecie i uzyskujemy dane.

Analiza 1: Korelacja Pearsona pomiędzy sprzedażą pieczywa a temperaturą

X Values	Y Values
12	702
14	872
18	902
11	872
13	702
14	852
19	972
10	982
9	912
8	952
2	1102
-2	1332
-8	1402
-2	1302
-1	1202
2	902
6	952
4	932
3	802
11	867



źródło: https://www.socscistatistics.com/tests/pearson/default2.aspx

Po wklejeniu danych do kalkulatora internetowego otrzymaliśmy informację, że poziom korelacji wynosi -0,75. To wysoki poziom współzależności. Czym niższa temperatura na ulicy tym większa sprzedaż pieczywa w sklepie. W analizie 1 widać, że spadek temperatury do poziomu 2°C pokazany w kolumnie „X Values" powoduje wzrost sprzedaży chleba powyżej tysiąca bochenków.

Porównywanie danych dyskretnych

Dane dyskretne to takie dane, których nie można zapisać za pomocą ułamków lub gdzie występuje mało unikatowych wartości. Danymi dyskretnymi są na przykład dni tygodnia. Jeżeli ponumerujemy dni tygodnia, to będziemy mieli 7 unikatowych wartości, gdzie poniedziałek to jeden, wtorek to dwa itd. Używając dyskretnych wartości nie można obliczyć korelacji z wartościami ciągłymi. Czyli nieprawidłowe jest obliczanie korelacji pomiędzy dniem tygodnia a wielkością sprzedaży.

Temperatura jest wartością ciągłą; gdybyśmy chcieli zrobić z niej wartość dyskretną musielibyśmy podzielić ją na zakresy temperatur, np. temperatura dodatnia (oznaczona jako 1) i ujemna (oznaczona jako 0). Dyskretny format mają numery dni miesiąca, tworzące zbiór liczb z zakresu od 1 do 31, dzień wolny i dzień pracujący, pora roku, słońce i brak słońca. Można przyjąć, że zmiennych dyskretnych jest więcej niż ciągłych.

Czerwona maszyna przynosi pecha. Kolor też jest informacją dyskretną, pech też, bo możemy przyjąć, że raz na kwartał zdarza się tam wypadek lub nie.

Skoro do pomiaru zależności zmiennych dyskretnych nie można użyć korelacji, w jaki sposób zbadać wzajemne powiązania danych? Z pomocą przychodzi test T.

Test -t

W naszym przypadku powinniśmy skorzystać z niesparowanego testu t. Test określa, czy istnieje różnica między dwiema teoretycznie niepowiązanymi grupami. Metoda porównuje średnią wartość z dwóch próbek. Jeżeli różnica średnich jest wystarczająco duża, zakłada się, że te dwie grupy różnią się między sobą. Możemy więc sprawdzić czy istnieje różnica w sprzedaży chleba przy temperaturze poniżej i powyżej 2°C.

Dzielimy więc nasz zbiór z „Analizy 1" na dwa zbiory opisujące ilość sprzedanego pieczywa. Pierwszy opisuje sprzedaż powyżej temperatury powyżej 2°C, drugi poniżej.

Wprowadzamy dane do kalkulatora internetowego testu-t w pozycji „Treatment 1 (*X*)" sprzedaż pieczywa w temperaturze powyżej 2°C. Kopiujemy do pozycji „Treatment 2 (*X*)" ilość sprzedanych bochenków w temperaturze poniżej 2°C. Kalkulator przeliczył istotną różnicę w średnich

Analiza 2. Test-t pomiędzy sprzedażą pieczywa a temperatura

Treatment 1 (*X*)	*Diff* (*X* - *M*)	*Sq. Diff* (*X* - *M*)^2
602	-271.25	73576.56
872	-1.25	1.56
902	28.75	826.56
872	-1.25	1.56
702	-171.25	29326.56
852	-21.25	451.56
972	98.75	9751.56
982	108.75	11826.56
952	78.75	6201.56
1102	228.75	52326.56
802	-71.25	5076.56
867	-6.25	39.06
	M: 873.25	SS: 189406.25

Treatment 2 (*X*)	*Diff* (*X* - *M*)	*Sq. Diff* (*X* - *M*)^2
1102	-21.67	469.44
1132	8.33	69.44
1302	178.33	31802.78
1102	-21.67	469.44
1202	78.33	6136.11
902	-221.67	49136.11
	M: 1123.67	SS: 88083.33

źródło: https://www.socscistatistics.com/tests/studentttest/default2.aspx

obu zbiorów, co oznacza, że granica temperatury 2°C ma istotny wpływ na zachowanie konsumentów.

Kalkulator wyświetlił następujący komunikat: The *t*-value is -3.80303. The *p*-value is .000781. The result is significant at $p < .05$.

Najważniejsza informacja w tym komunikacie mówi, że wartość p-value wynosi 0.000781 czyli jest niższa od progu 0.05. To oznacza, że próbki różnią się istotnie.

Test chi-kwadrat

Jest najbardziej znanym testem dla sprawdzenia zależności par dyskretnych. Innymi słowy odpowiada na pytanie: czy np. osoby starsze częściej kupują chleb, dzieci i osoby dorosłe odwiedzają sklep w innej częstotliwości.

Taka analiza może nam się przydać, ponieważ możemy zorganizować w środy darmową kawę dla seniora lub inną akcję, która pomoże w pogłębieniu relacji z określoną grupą mieszkańców. Aby zainwestować zasoby i całą energię (której zawsze jest za mało) musimy mieć pewność, że nasze przypuszczenia nie są „wierzeniami ludów pierwotnych".

Załóżmy, że mieszkańców mamy podzielonych na 3 grupy wiekowe: „dzieci i młodzież", „dorośli", „osoby 60+". Sprzedawczynie w sklepie dyskretnie w portalu sprzedażowym odnotowują informację o wieku kupującego. Pod koniec miesiąca mamy zebraną bazę struktury wiekowej kupujących.

Tabela 1: Liczba kupujących według kategorii wiekowych i dni tygodnia.

	Pn	Wt	Śr	Cz	Pi	So
Dzieci i młodzież	334	334	231	233	543	432
Dorośli	456	345	312	345	343	324
Osoby starsze	454	347	421	345	532	213

Dane zebrane w trakcie badania możemy wpisać do kalkulatora internetowego.

Analiza 3. Test chi-kwadrat dla kategorii wiekowych i dni tygodnia

	Gr 1	Gr 2	Gr 3	Gr4	Gr 5	Gr 6	Gr
Data 1:	334	334	231	233	543	432	
Data 2:	456	345	312	345	343	324	
Data 3:	454	347	421	345	532	213	
Data 4:							

Degree of Freedom(df)	=	10
Chi square test(χ^2)	=	224.008
P value	=	0

źródło: https://drr.ikcest.org/app/s3110

Ponieważ wartość „P value" wynosi zero (według moich obliczeń wynosi: 1.5473873958548322e-42), przyjmujemy, że istnieją istotne statystycznie różnice w odwiedzaniu sklepu pomiędzy trzema kategoriami wiekowymi w poszczególnych dniach tygodnia.

Oznacza to, że warto dalej badać zjawisko.

Podsumowanie

Testy statystyczne robi się po to, aby dowiedzieć się czy obserwowane zjawisko jest istotne czy jest to tylko nasze urojenie, tzw. „miejska legenda". Testy stosuje się, aby wybrać zmienne, takie jak dzień tygodnia, poziom opadów czy sezon do modelu regresji. Taki model jest łatwy w opracowaniu w bezpłatnych kalkulatorach internetowych. Może on z dużym prawdopodobieństwem przewidzieć przyszłą sprzedaż, ceny surowców lub poziom przyszłej efektywności kampanii reklamowej.

Wojciech Moszczyński

Wojciech Moszczyński – ekspert z zakresu optymalizacji matematycznej oraz modelowania predykcyjnego. Od lat zaangażowany w popularyzację metod ekonometrycznych w środowiskach biznesowych. Specjalizuje się w optymalizacji procesów sprzedażowych, produkcyjnych oraz logistycznych. Przez 15 lat pracował jako ekspert finansowy ze specjalizacją w obszarze kontrolingu i rachunkowości zarządczej. Od 10 lat pracuje jako analityk danych (*data scientist*). Absolwent katedry Ekonometrii i Statystyki Uniwersytetu Mikołaja Kopernika w Toruniu. Obecnie zatrudniony jako Senior Data Scientist w polskiej firmie Unity Group.