

ANOM - Analiza šrednich.

Kamila Komar i Marta Sommer

26.05.14

Plan prezentacji

1 Wprowadzenie

Plan prezentacji

- 1 Wprowadzenie
- 2 ANOM jednoczynnikowa

Plan prezentacji

- 1 Wprowadzenie
- 2 ANOM jednoczynnikowa
- 3 ANOM dla zadanych specyfikacji

Plan prezentacji

- 1 Wprowadzenie
- 2 ANOM jednoczynnikowa
- 3 ANOM dla zadanych specyfikacji
- 4 ANOM wieloczynnikowa

W wielu sytuacjach, potrzebujemy zmierzyć się z problemem porównywania średnich w więcej niż dwóch grupach. Do tego właśnie służy ANOM-analiza średnich. Dzięki testowi wprowadzonemu przez E. R. Ott możemy sprawdzić czy i które ze zbioru c -średnich różnią się od średniej ogólnej. Zamiast patrzeć na górne i dolne granice przedziału ufności, patrzymy które z c -grup średnich wychodzą poza górną i dolną linię decyzji. Każda średnia powyżej linii decyzyjnej górnej, jest uważana za znacznie wyższą od ogólnej. Podobnie, średnie leżące poniżej dolnej granicy są uznawane za znacznie mniejsze od ogólnej średniej.

W przypadku testowania pojedynczego odstępstwa $\overline{X_1} = \overline{X_2}$ korzystamy z testu istotności:

$$t = \frac{\overline{x_1} - \overline{x_2}}{S_{\overline{X_1} - \overline{X_2}}},$$

gdzie:

$\overline{X_1}, \overline{X_2}$ - średnie z prób

n_1, n_2 - liczności grup

Zatem możemy powiedzieć, że dwie próbki różnią się gdy $t > t_\alpha$, gdzie t_α poziom krytyczny na poziomie istotności α .

A CO GDY MAMY WIĘCEJ NIŻ DWIE GRUPY?

Czyli chcemy testować:

$$H_0 : \overline{X}_i = \overline{\overline{X}},$$

gdzie $\overline{\overline{X}}$ - średnia ogólna.

Ogólny pomysł to wyznaczenie linii decyzyjnych.

Dolną i górną (LDL i UDL) granicę decyzji wyznaczamy ze wzoru:

$$\bar{\bar{X}} \pm h_{\alpha,k,v} S \sqrt{\frac{(k-1)}{kn}},$$

gdzie:

k – liczba rozważanych próbek

n – liczność grup

$\bar{\bar{X}}$ – ogólna średnia

S – estymator wariancji.

$h_{\alpha,k,v}$ – krytyczna wartość statystyki Nelsona

v – liczba stopni swobody związana z S ($= k \cdot (n - 1)$)

Można udowodnić, że $S \sqrt{\frac{(k-1)}{kn}}$ to estymator wariancji $\bar{X}_i - \bar{\bar{X}}$.

Rozważmy dane opisujące wpływ temperatury na wydajność procesu:

	Temperatura		
Dzień	250°	300°	350°
Pn	2,4	2,6	3,2
Wt	2,7	2,4	3,0
Śr	2,2	2,8	3,1
Czw	2,5	2,5	2,8
Pt	2,0	2,2	2,5
Pn	2,5	2,7	2,9
Wt	2,8	2,3	3,1
Śr	2,9	3,1	3,4
Czw	2,4	2,9	3,2
Pt	2,1	2,2	2,6

$$n = 10, \quad k = 3, \quad \bar{\bar{X}} = 2.67$$

Tak jak to robiliśmy w ANOVA możemy wyznaczyć estymator odchylenia standardowego:

$$S = \hat{\sigma} = \sqrt{0.0867}$$

$$\nu = 3 \cdot (10 - 1) = 27$$

$$\alpha = 0.05$$

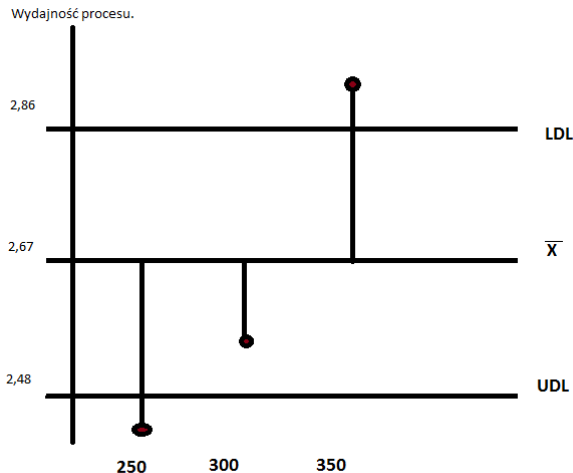
Z tablic dla statystyki Nelsona odczytujemy wartość $h_{0.05,3,27} = 2.485$
Podstawmy wszystkie wartości do wzoru:

$$\bar{\bar{X}} \pm h_{\alpha,k,\nu} S \sqrt{\frac{(k-1)}{kn}},$$

W wyniku otrzymujemy:

$$LDL = 2.48 \text{ i } UDL = 2.86$$

Graficznie, można to przedstawić w następujący sposób:



Widać, że wydajność procesu spada przy 250° i jest wyższa niż ogólna średnia przy 350°.

Założenia, dla tej metody są identyczne jak w przypadku ANOVY:

- niezależność k próbek
- normalność rozkładu
- jednakowe wariancje

Rozważmy dane opisujące firmę (przemysł rolny) zatrudniającą osoby obsługujące kombajny. Jesteśmy zainteresowani oceną względnej wydajności każdego z pracowników. Rejestrujemy procent upraw nie spełniających specyfikacji, dla każdego pracownika. Chcemy zidentyfikować pracownika, który ma najgorsze wyniki, aby go przeszkolić na nowo lub zwolnić.

Firma zatrudnia 20 pracowników.

Numer pracownika	Niezgodność upraw (%) [n=1000]
1	1,4
2	2,6
3	1,0
4	3,1
5	2,9
6	5,1
7	2,4
8	4,1
9	1,1
10	2,1
11	2,0
12	2,6
13	3,1
14	2,7
15	...

W przypadku takich danych linie decyzyjne wyznaczamy ze wzoru:

$$\bar{p} \pm h_{\alpha,k,\infty} \cdot S_p \cdot \sqrt{\frac{k-1}{k}},$$

gdzie:

\bar{p} – średnia ze wszystkich proporcji

$S_p = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ – estymator σ_p

k – liczba grup

W przypadku wspomnianych danych mamy:

$$\bar{p} = 0.02775$$

$$S_p = \sqrt{\frac{0.02775 \cdot 0.97225}{1000}} = 0.0052$$

$$k = 20$$

Z tabeli wartości statystyki Nelsona odczytujemy wartość:

$$h_{0.05,20,\infty} = 3.02.$$

W praktyce częściej przyjmuje się poziom istotności $\alpha = 0.01$ i wtedy

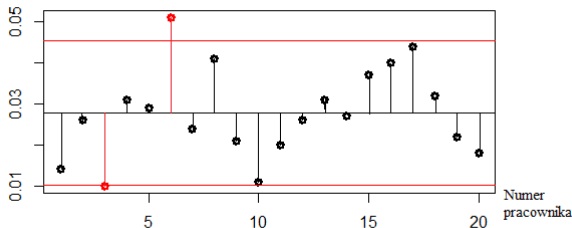
$$h_{0.01,20,\infty} = 3.48.$$

W ten sposób otrzymaliśmy

$$LDL = 0,0101 \text{ i } UDL = 0.0454$$

Nasze wyniki przedstawione graficznie:

Proporcje złych upraw



Z powyższej postaci graficznej od razu widać, że mamy 11 obserwacji powyżej średniej i 9 poniżej. Ponadto wynik obserwowany dla pracownika o numerze 6 przekracza UDL. Pracownik nr. 3 ma wynik leżący poniżej LDL oraz pracownik 9 bardzo zbliża się do niej.

ANOM możemy używać, gdy mamy odczynienia z danymi dyskretnymi, które powinny być opisywane rozkładem Poissona. Linie decyzyjne w tym przypadku wyznacza się ze wzoru:

$$\bar{c} \pm h_{\alpha,k,\infty} \cdot \sqrt{\bar{c}} \cdot \sqrt{\frac{k-1}{k}},$$

gdzie:

\bar{c} - średnia z k -zliczeń

Możliwość użycia ANOM w tym przypadku opiera się na założeniu aproksymacji rozkładu Poissona do rozkładu normalnego. Jeśli dane nie spełniają założeń, należy dokonać ich transformacji.

ANOM przy zadanych standardach

Dla proporcji:

$$p \pm h_{\alpha,k,\infty} \cdot \sqrt{\frac{p(1-p)}{n}}$$

Dla danych dyskretnych:

$$c \pm h_{\alpha,k,\infty} \cdot \sqrt{c}$$

ANOM przy zadanych standardach

Dla danych liczbowych (σ i μ znane):

$$\mu \pm h_{\alpha,k,\infty} \cdot \frac{\sigma}{\sqrt{n}}$$

Dla danych liczbowych (μ jest nieznane, a σ jest znane):

$$\bar{\bar{x}} \pm h_{\alpha,k,\infty} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{k-1}{k}}$$

Dla danych liczbowych (μ jest znane, a σ jest nieznane):

$$\mu \pm h_{\alpha,k,\nu} \cdot \frac{S}{\sqrt{n}}$$

ANOM wieloczynnikowa

Efekt A i B.

Linie kontrolne dla efektów A i B wyznaczamy standardowo ze wzoru:

$$\bar{\bar{x}} \pm h_{\alpha,k,\infty} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{k-1}{k}}$$

Ale należałoby też uwzględnić integracje!

Interakcje

Interakcje wyznaczamy z następującego wzoru:

$$\bar{L} = \frac{1}{2}(\overline{A_1 B_1} + \overline{A_2 B_2})$$

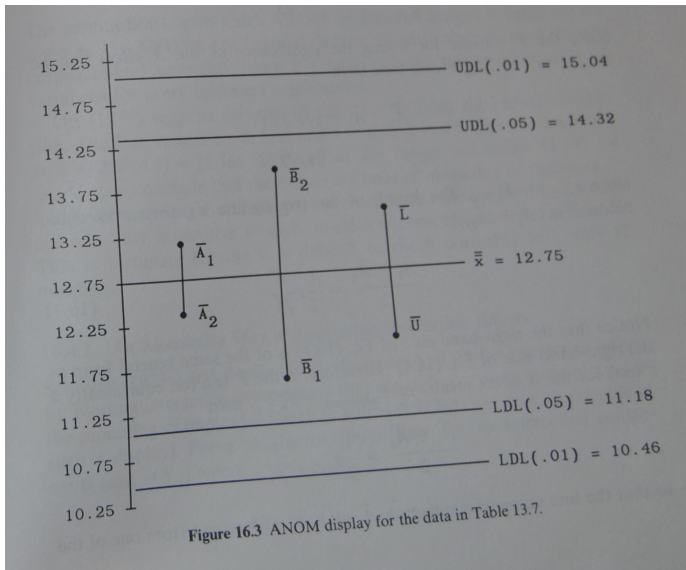
$$\bar{U} = \frac{1}{2}(\overline{A_1 B_2} + \overline{A_2 B_1})$$

Wtedy:

$$AB = \bar{L} - \bar{U}$$

Linie decyzyjne dla AB wyznaczane są z tego samego wzoru, co dla A i B.

Rysunek



Inny sposób wyznaczania interakcji

$$AB_{ij} = \overline{AB_{ij}} - \overline{A_i} - \overline{B_j} + \overline{\bar{x}}$$

Problemy:

- inna skala rysunku niż dla średnich w grupach
- nie znamy wartości h_α dla takiej statystyki

ANOM wieloczynnikowa, przy wielu poziomach czynnika

Table 16.2 Length of Steel Bars (Coded Data)^a

Time	Heat Treatment							
	W				L			
	Machine				Machine			
	A	B	C	D	A	B	C	D
1	6	7	1	6	4	6	-1	4
	9	9	2	6	6	5	0	5
	1	5	0	7	0	3	0	5
	3	5	4	3	1	4	1	4
2	6	8	3	7	3	6	2	9
	3	7	2	9	1	4	0	4
	1	4	1	11	1	1	-1	6
	-1	8	0	6	-2	3	1	3
3	5	10	-1	10	6	8	0	4
	4	11	2	5	0	7	-2	3
	9	6	6	4	3	10	4	7
	6	4	1	8	7	0	-4	0

^aThese data originally appeared in Baten (1956). © 1956 American Society for Quality Control. Reprinted by permission.

ANOM wieloczynnikowa, przy wielu poziomach czynnika

Stosujemy po prostu do każdego czynnika zwykłą ANOM.

Tak więc:

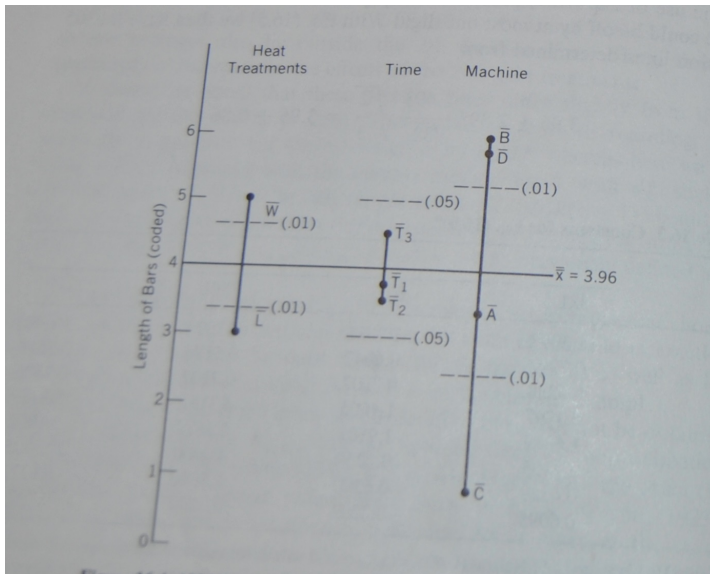
$$\overline{T}_1 = 3,78, \quad \overline{T}_2 = 3,625, \quad \overline{T}_3 = 4,47$$

$$\overline{W} = 4,98, \quad \overline{L} = 2,94$$

$$\overline{A} = 3,42, \quad \overline{B} = 5,875, \quad \overline{C} = 0,875, \quad \overline{D} = 5,67$$

$$\overline{\overline{x}} = 3,96$$

ANOM wieloczynnikowa, przy wielu poziomach czynnika



BIBLIOGRAFIA



Thomas P. Ryan *Statistical Methods for Quality Improvement*, WILEY (1989), rozdział 16.

;)

DZIĘKUJEMY!