Rozłączność przedziałów ufności jako kryterium weryfikacji hipotez statystycznych

Tomasz Żółtak

Instytut Badań Edukacyjnych

24.09.2018

Wprowadzenie

Motywacja

- Namysł nad sposobem wizualizacji *trzyletnich wskaźników EWD*.
    - Dwuwymiarowe obszary ufności wykorzystywane do charakteryzowania poszczególnych szkół, ale również do wnioskowania o istotności różnic między szkołami lub istotności zmian w czasie.
    - Bardzo niszowy temat.
- Problem ogólny:
    - Jako badacze stawiamy i weryfikujemy statystycznie hipotezy korzystając z procedury Neymana-Pearsona.
    - Ale na wykresach pokazujemy przedziały ufności.
      W szczególności na wykresach obrazujących efekty brzegowe w modelach regresji.

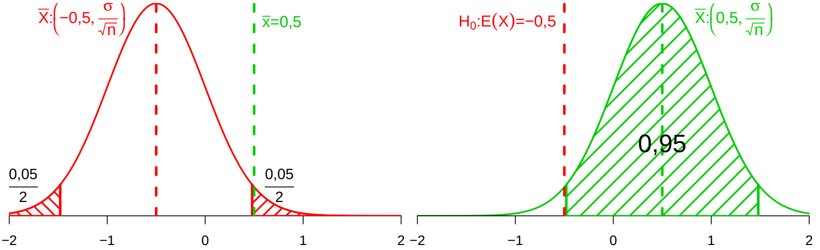
Powszechnie znany związek

W sytuacji gdy:

- Możemy przyjąć, że rozkład statystyki z próby jest symetryczny (np. normalny).
- Wariancja rozkładu statystyki z próby nie wynika z hipotezy zerowej.

Jeśli przedział ufności dla parametru wyznaczony na poziomie ufności $\gamma = 1 - \alpha$ nie obejmuje wartości *c*,
to w dwustronnym teście na poziomie istotności α odrzucona zostanie H_0 głosząca, że populacyjna wartość parametru jest równa *c*.

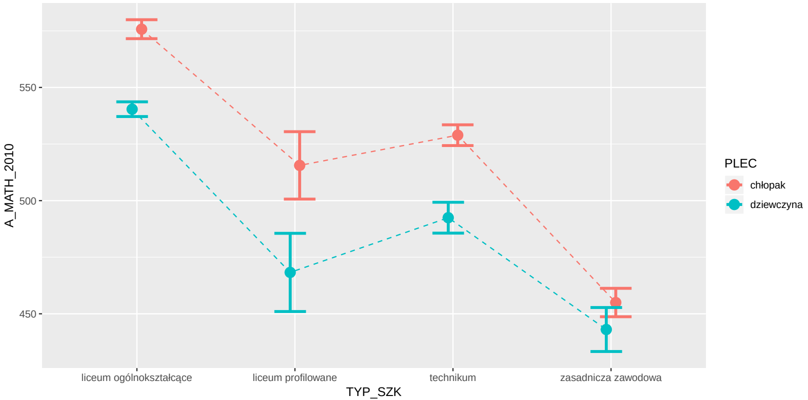
Przykładowo - weryfikacji hipotezy dot. średniej populacyjnej:



Często występująca pokusa

Wykres efektów brzegowych (*marginal effects at the mean*) z 95% PU:

W modelu regresji zmienną niezależną był dodatkowo wynik testu inteligencji. Przy obliczaniu prezentowanych wartości ustalono jej wartość w średniej dla analizowanej zbiorowości.



Czy płeć różnicuje osiągnięcia uczniów z matematyki w zasadniczych szkołach zawodowych?

Czy różnica widoczna na wykresie jest statystycznie istotna?

Model regresji

Osiągnięcia z matematyki 3615 uczniów polskich szkół pogimnazjalnych, mierzone testem PISA z 2009 r., przewidywane są ze względu na:

- wynik testu inteligencji ogólnej (zmienna ciągła *RAVEN WYN*),
- płeć (zmienna binarna *PLECchlopak*: 0-dziewczyna, 1-chłopak),
- typ szkoły (trzy zmienne binarne: *TYP_SZK...*),
- interakcje płci i typu szkoły (trzy zmienne binarne *PLECchlopak:TYP_SZK...*).

parametr	wartość	bł. std.	t	ist.
b_0 stała	194,55	7,77	25,04	0,00
b_1 RAVEN_WYN	7,14	0,15	47,48	0,00
b_2 PLECchłopak	35,34	2,64	13,37	0,00
b_3 TYP_SZKliceum profilowane	-72,12	8,98	-8,03	0,00
b_4 TYP_SZKtechnikum	-47,94	3,88	-12,37	0,00
b_5 TYP_SZKzasadnicza zawodowa	-97,34	5,31	-18,34	0,00
b_6 PLECchłopak:TYP_SZKliceum profilowane	11,95	11,92	1,00	0,32
b_7 PLECchłopak:TYP_SZKtechnikum	1,12	4,94	0,23	0,82
b_8 PLECchłopak:TYP_SZKzasadnicza zawodowa	-23,42	6,22	-3,77	0,00

Co chcemy obliczyć?

Na wykresie widzimy:

- Średnie wyniki dziewczyn w ZSZ (przy ustalnej wartości *RAVEN_WYN*): $b_{kZSZ} = b_0 + b_1 E(RAVEN\_WYN) + b_5$
- Średnie wyniki chłopaków w ZSZ (przy ustalnej wartości *RAVEN_WYN*): $b_{mZSZ} = b_0 + b_1 E(RAVEN\_WYN) + b_2 + b_5 + b_8$

Interesującą nas różnicę między płciami w ZSZ opisuje:

- $r_{ZSZ} = b_{mZSZ} - b_{kZSZ} = b_2 + b_8$

parametr	wartość	bł. std.	t	ist.
b_0 stała	194,55	7,77	25,04	0,00
b_1 RAVEN_WYN	7,14	0,15	47,48	0,00
b_2 PLECchłopak	35,34	2,64	13,37	0,00
b_3 TYP_SZKliceum profilowane	-72,12	8,98	-8,03	0,00
b_4 TYP_SZKtechnikum	-47,94	3,88	-12,37	0,00
b_5 TYP_SZKzasadnicza zawodowa	-97,34	5,31	-18,34	0,00
b_6 PLECchłopak:TYP_SZKliceum profilowane	11,95	11,92	1,00	0,32
b_7 PLECchłopak:TYP_SZKtechnikum	1,12	4,94	0,23	0,82
b_8 PLECchłopak:TYP_SZKzasadnicza zawodowa	-23,42	6,22	-3,77	0,00

Jakie jest oszacowanie błędu standardowego tej statystyki?

Jak dobrze wiadomo: $D^2(X \pm Y) = D^2(X) + D^2(Y) \pm 2Cov(X,Y)$

Na podstawie macierzy kowariancji parametrów modelu:

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
b_0	60,38	-1,14	-1,11	-8,96	-7,03	-15,42	3,75	2,18	2,96
b_1	-1,14	0,02	-0,03	0,12	0,09	0,25	-0,02	0,01	-0,01
b_2	-1,11	-0,03	6,99	2,49	2,54	2,31	-6,92	-6,96	-6,94
b_3	-8,96	0,12	2,49	80,72	3,13	4,05	-80,15	-2,60	-2,69
b_4	-7,03	0,09	2,54	3,13	15,02	3,62	-2,74	-14,65	-2,68
b_5	-15,42	0,25	2,31	4,05	3,62	28,18	-2,90	-2,55	-25,42
b_6	3,75	-0,02	-6,92	-80,15	-2,74	-2,90	142,05	6,94	6,95
b_7	2,18	0,01	-6,96	-2,60	-14,65	-2,55	6,94	24,43	6,95
b_8	2,96	-0,01	-6,94	-2,69	-2,68	-25,42	6,95	6,95	38,63

możemy więc obliczyć: $D(r_{ZSZ}) = \sqrt{D^2(b_2) + D^2(b_8) + 2Cov(b_2, b_8)} = \sqrt{D^2(b_{mZSZ}) + D^2(b_{kZSZ}) - 2Cov(b_{mZSZ}, b_{kZSZ})}$

parametr	wartość	bł. std.	(ko)wariancja	t	ist.
r_{ZSZ}	11,92	5,63	31,74	2,12	0,03
b_{kZSZ}	443,06	4,96	24,56		
b_{mZSZ}	454,98	3,20	10,24		
$2Cov(b_{kZSZ}, b_{mZSZ})$			3,07		

Czy to w ogóle kiedyś działa?

- Aby reguła *rozłącznych przedziałów ufności* działała, suma szerokość PU dla dwóch efektów brzegowych musiałby być równa szerokości PU dla różnicy, co jest równoważne:
  $D(r_{ZSZ}) = D(b_{kZSZ}) + D(b_{mZSZ})$
- Taka równość zachodzi **w dwóch specyficznych przypadkach**:
  1. $D(b_{kZSZ}) = 0$ lub $D(b_{mZSZ}) = 0$
     Sytuacja tożsama z przyrównywaniem do znanej stałej *c*.
     Ten warunek będzie **w przybliżeniu spełniony**, gdy:
     $D(b_{kZSZ}) \gg D(b_{mZSZ})$ lub $D(b_{kZSZ}) \ll D(b_{mZSZ})$
     Ale i to jest niezwykle nietypowa sytuacja.
  2. $cor(b_{kZSZ}, b_{mZSZ}) = -1$
- Ale w ogólności **nie ma podstaw, by tego oczekiwać!**
- Generalnie $D(r_{ZSZ}) < D(b_{kZSZ}) + D(b_{mZSZ})$, a rozbieżność jest tym większa, im bardziej:
  - zbliżone są do siebie wartości błędów standardowych porównywanych efektów brzegowych,
  - korelacja porównywanych efektów brzegowych odbiega od -1.

Jak sobie radzić?

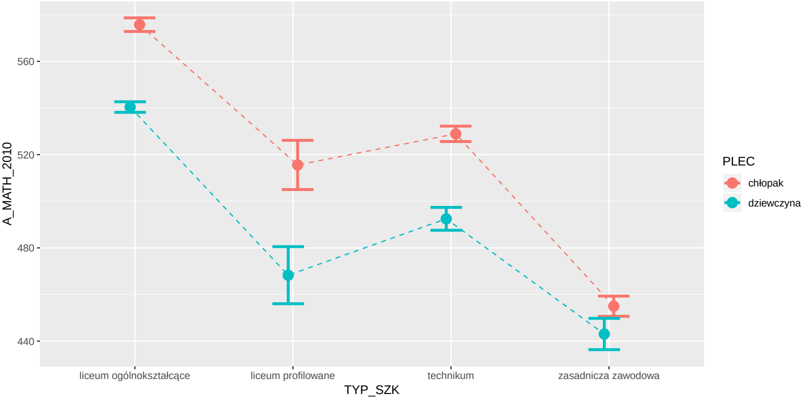
Kilka możliwych rozwiązań

- Nie rysować wykresów, gdy mówimy o weryfikacji hipotez dotyczących różnicy wartości (kombinacji liniowych) parametrów modelu.

Kilka możliwych rozwiązań

- Nie rysować wykresów, gdy mówimy o weryfikacji hipotez dotyczących różnicy wartości (kombinacji liniowych) parametrów modelu.
- *Przeskalować* (przemnożyć) szerokość rysowanych przedziałów ufności o czynnik $\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$

Przeskalowanie PU - ilustracja



TYP_SZK	$\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$
liceum ogólnokształcące	0,70
liceum profilowane	0,71
technikum	0,72
zasadnicza zawodowa	0,69

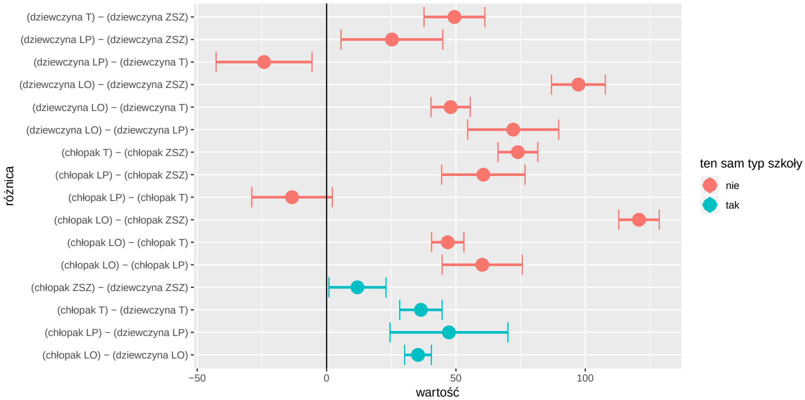
Kilka możliwych rozwiązań

- Nie rysować wykresów, gdy mówimy o weryfikacji hipotez dotyczących różnicy wartości (kombinacji liniowych) parametrów modelu.
- *Przeskalować* (przemnożyć) szerokość rysowanych przedziałów ufności o czynnik $\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$
  - **Problem 1.:** na jednym wykresie każdy efekt brzegowy możemy porównać tylko z jednym innym efektem brzegowym.
  - **Problem 2.:** łatwo może wprowadzić odbiorcę w błąd - wykres wygląda jak typowy wykres z PU dla efektów brzegowych, ale nim nie jest.

Kilka możliwych rozwiązań

- Nie rysować wykresów, gdy mówimy o weryfikacji hipotez dotyczących różnicy wartości (kombinacji liniowych) parametrów modelu.
- *Przeskalować* (przemnożyć) szerokość rysowanych przedziałów ufności o czynnik $\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$
    - Problem 1.: na jednym wykresie każdy efekt brzegowy możemy porównać tylko z jednym innym efektem brzegowym.
    - Problem 2.: łatwo może wprowadzić odbiorcę w błąd - wygląda jak typowy wykres z PU dla efektów brzegowych, ale nim nie jest.
- Narysować wykres z testowanymi różnicami i PU dla tych różnic.

Wykres różnic parami z PU



Na wykresie pominięto 12 par efektów brzegowych, w których jednocześnie różniły się od siebie wartości zmiennej *PLEC* i zmiennej *TYP_SZK*.

Kilka możliwych rozwiązań

- Nie rysować wykresów, gdy mówimy o weryfikacji hipotez dotyczących różnicy wartości (kombinacji liniowych) parametrów modelu.
- *Przeskalować* (przemnożyć) szerokość rysowanych przedziałów ufności o czynnik $\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$
  - Problem 1.: na jednym wykresie każdy efekt brzegowy możemy porównać tylko z jednym innym efektem brzegowym.
  - Problem 2.: łatwo może wprowadzić odbiorcę w błąd - wygląda jak typowy wykres z PU dla efektów brzegowych, ale nim nie jest.
- Narysować wykres z testowanymi różnicami i PU dla tych różnic.
  - I to rozwiązanie polecam.

Modele z nieliniową funkcją łączącą

- *Przeskalować* (przemnożyć) szerokość rysowanych przedziałów ufności o czynnik $\frac{D(r_\bullet)}{D(b_{k\bullet})+D(b_{m\bullet})}$
  - Problem 1.: na jednym wykresie każdy efekt brzegowy możemy porównać tylko z jednym innym efektem brzegowym.
  - Problem 2.: łatwo może wprowadzić odbiorcę w błąd - wygląda jak typowy wykres z PU dla efektów brzegowych, ale nim nie jest.
- Narysować wykres z testowanymi różnicami i PU dla tych różnic.
  - I to rozwiązanie polecam.

W regresjach z nieliniową funkcją łączącą (np. logistycznej, probitowej, Poissona, ujemnej dwumianowej) żadne z powyższych rozwiązań **nie pozwala przygotować wykres, na którym efekty byłyby przedstawione na skali zmiennej zależnej** (bez przyjęcia dodatkowych arbitralnych założeń).

Koniec

Dziękuję za uwagę!

Tomasz Żółtak
t.zoltak@ibe.edu.pl