WIADOMOŚCI STATYSTYCZNE

PL ISSN 0043-518X
e-ISSN 2543-8476

Rok LXI
11 (666) 2016, 7—12

STUDIA METODOLOGICZNE

Wojciech MŁOCEK

Uwagi do artykułu pt. „O złej radzie dotyczącej testu F Snedecora"

Streszczenie. *W opracowaniu (Drapella, 2016) autor zawarł tezę o błędnym sposobie obliczania statystyki testowej w teście Fishera-Snedecora (FS). Sposób ten jest opisywany powszechnie w literaturze, a przykłady, które przytacza autor, pochodzą z podręczników Brandta (1998), Starzyńskiej (2002) i tablic statystycznych Zielińskiego (1972). Tezę poparto eksperymentem numerycznym oraz rozważaniami analitycznymi. W niniejszym artykule wskazano błąd w dyskutowanym opracowaniu (Drapella, 2016), tym samym neguje się zawartą w nim tezę oraz pokazuje, że wspomniana literatura nie zawiera błędnych informacji związanych z testem FS.*

Słowa kluczowe: test Fishera-Snedecora.

TEST FISHERA-SNEDECORA

Test Fishera-Snedecora (FS), który jest głównym tematem niniejszej pracy, jest opisany w prawie każdym podręczniku obejmującym podstawowy kurs statystyki. Test ten opiera się na statystyce F, którą sformułował Fisher (1925), a opisał Snedecor (1934) i dotyczy równości wariancji dwóch niezależnych zmiennych losowych mających rozkłady normalne.

Niech $X_1 \sim N(m_1, \sigma_1)$, $X_2 \sim N(m_2, \sigma_2)$ będą niezależnymi zmiennymi losowymi, których rozkłady są normalne o nieznanych parametrach. Załóżmy też, że $X_{11}, X_{12}, ..., X_{1n_1}$ oraz $X_{21}, X_{22}, ..., X_{2n_2}$ są ciągami niezależnych zmiennych losowych o takim samym rozkładzie, jak rozkład zmiennych X_1, X_2 odpowiednio. Ponadto, przyjmijmy że:

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \quad \overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$$

oraz

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(X_{2i} - \overline{X}_2)^2.$$

Wartości powyższych statystyk oznaczamy małymi literami, czyli mamy odpowiednio $\overline{x}_1$, $\overline{x}_2$, s_1^2, s_2^2.

Stawiamy hipotezę zerową, że wariancje zmiennych X_1, X_2 są równe:

$$H_0 : \sigma_1^2 = \sigma_2^2.$$

Hipotezą alternatywną może być jedna z trzech hipotez:

(a) $H_1 : \sigma_1^2 \neq \sigma_2^2$,

(b) $H_1 : \sigma_1^2 > \sigma_2^2$,

(c) $H_1 : \sigma_1^2 < \sigma_2^2$.

Przy prawdziwości H_0 statystyka testowa jest następująca:

$$F = \frac{S_1^2}{S_2^2} \tag{1}$$

i ma ona rozkład FS z n_1–1 oraz n_2–1 stopniami swobody ($F \sim F(n_1-1,\ n_2-1)$). Wartość statystyki testowej, wyznaczona na podstawie konkretnej próby prostej, jest zatem określona wzorem:

$$f = \frac{s_1^2}{s_2^2}. \tag{2}$$

Jeśli przez f_{p, r_1, r_2}, oznaczymy kwantyl rzędu p, a przez F_{r_1, r_2} dystrybuantę rozkładu $F(r_1, r_2)$, to przy poziomie istotności α zbiory krytyczne K oraz p-wartości (krytyczne poziomy istotności) mają postać, odpowiednio:

(a) $K = (0,\ f_{\frac{\alpha}{2}, n_1-1, n_2-1}] \cup [f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}, +\infty)$,

$p = 2 \cdot \min\{1 - F_{n_1-1, n_2-1}(f),\ F_{n_1-1, n_2-1}(f)\}$,

(b) $K=[f_{1-\alpha,n_1-1,n_2-1}+\infty)$, $p=1-F_{n_1-1,n_2-1}(f)$,

(c) $K=(0,\ f_{\alpha,n_1-1,n_2-1}]$, $p=F_{n_1-1,n_2-1}(f)$.

PRAKTYCZNE ZASTOSOWANIE TESTU FISHERA-SNEDECORA

Zdecydowana większość podręczników, z których korzystamy w nauczaniu statystyki dzisiaj powstawała w czasach, gdy tablice statystyczne stanowiły dodatki do nich. Co więcej, były one dostępne zazwyczaj tylko w takiej formie. Zrozumiałym jest, że w takich tablicach nie można było zawrzeć wszystkiego. Wszędzie tam, gdzie było to możliwe starano się podawać tylko minimalną ilość informacji. Klasycznym przykładem są tablice FS, które są niezbędne do stosowania testu FS w praktyce. Otóż wystarczy znać tylko kwantyle rzędu 0,975 oraz 0,95 rozkładu FS, by stosować test FS na poziomie istotności 0,05. Przedstawimy teraz odpowiedź na pytanie, dlaczego w tej sytuacji nie potrzebujemy dodatkowo kwantyli rzędu 0,025 oraz 0,05. Jest to jednocześnie odpowiedź na pytanie, w którym miejscu rozumowanie w przywołanym artykule (Drapella, 2016) jest błędne.

Sposób, w jaki ponumerujemy populacje (nr 1, nr 2) jest zawsze sprawą umowną. Stąd nie może to mieć i nie ma wpływu na wynik żadnego testu statystycznego. W przypadku testu FS przyjmujemy taką numerację populacji, by $s_1^2 \geq s_2^2$. Możemy przyjąć zasadę, że numeracji dokonujemy dopiero po obliczeniu wariancji z obu prób. Tym samym wartość naszej statystyki testowej z próby, która jest określona wzorem (2) będzie zawsze spełniała nierówność $f \geq 1$. W tej sytuacji jedynymi rozsądnymi hipotezami alternatywnymi są postaci (a) i (b), które wyczerpują w pełni wszystkie możliwości. Przyjmując bowiem hipotezę alternatywną postaci (c), z góry wiadomo, przy „rozsądnym poziomie istotności", że nie ma podstaw do jej odrzucenia. Co więcej, dla hipotezy alternatywnej postaci (a) zbiór krytyczny możemy ograniczyć do zbioru $K=[f_{1-\frac{\alpha}{2},n_1-1,n_2-1},+\infty)$, gdyż wartość statystyki testowej z próby nie trafi na pewno do przedziału $(0,\ f_{\frac{\alpha}{2},n_1-1,n_2-1}]$.

Dodajmy, że obecnie powszechnie dostępne programy statystyczne (np. R, Statistica czy Excel) mają zdefiniowane funkcje służące do wyznaczania kwantyli dużej liczby rozkładów, w tym rozkładu FS. Opisaną zatem metodę można uznać za „przestarzałą", co nie oznacza oczywiście, że jest ona nieprawdziwa. Warto również wspomnieć, że we współczesnych pracach naukowych, w których stosuje się testy statystyczne, zamiast wartości statystyki testowej z próbki i postaci zbioru krytycznego podaje się zazwyczaj samą p-wartość.

PRACA PT. „O ZŁEJ RADZIE DOTYCZĄCEJ TESTU F SNEDECORA"

W pracy (Drapella, 2016) tytułową radę autor odnosi do sposobu numeracji populacji, tak by $s_1^2 \geq s_2^2$. W powyższym rozdziale pokazano, że nie ma to wpływu na test FS. Powstaje zatem pytanie, dlaczego eksperyment numeryczny w przedmiotowym opracowaniu pokazał, że ma to wpływ na błąd pierwszego rodzaju. Sam eksperyment został przeprowadzony poprawnie. Błąd tkwi w przyjętej granicy obszaru dopuszczalnego w schemacie S_2. Skoro autor, w obu schematach (S_1, S_2), jako hipotezę alternatywną rozważa, że wariancje są różne, to w schemacie S_2 powinno być $F_g^{kr} = 1{,}861$ zamiast $F_g^{kr} = 1{,}620$. Przy $F_g^{kr} = 1{,}620$ w schemacie S_2 autor weryfikuje tak naprawdę hipotezę przy $H_1 : \sigma_1^2 > \sigma_2^2$ lub $H_1 : \sigma_1^2 < \sigma_2^2$, dlatego eksperyment wskazał błąd pierwszego rodzaju oscylujący wokół wartości 2α.

Dyskutowany artykuł (Drapella, 2016), poza eksperymentem numerycznym, zawierał również rozważania analityczne. Autor powiela w nich błąd polegający na wzięciu niewłaściwej granicy obszaru dopuszczalnego w schemacie S_2, który w rezultacie daje błędną tezę. Co więcej, rozważania te dotyczą tylko szczególnego przypadku $(F(r_1, r_2)$ przy $r_1 = r_2 = 29)$. Przypomnijmy (Zieliński, 1972, s. 20), że jeśli $X \sim F(r_1, r_2)$, to $1/X \sim F(r_2, r_1)$. Weźmy $X \sim F(r_1, r_2)$ i przyjmijmy, że $Y = \max(X, 1/X)$, wtedy dystrybuantę zmiennej losowej Y możemy określić wzorem $F_Y(y) = F_X(y) - F_X(1/y)$. Jeśli dodatkowo założymy, że $r_1 = r_2$, to $F_Y(y) = 2\,F_X(y) - 1$. Gdy w schemacie S_2 przyjmiemy, że statystyką testową jest Y, to granicą dopuszczalną jest kwantyl rzędu $1-\alpha$ zmiennej Y. W ogólnym przypadku $(r_1 \neq r_2)$ nie jest to prawdą, dlatego wykr. 3 w omawianej tu pracy (Drapella, 2016) może wprowadzać Czytelnika w błąd.

Uwagi końcowe

Podręczniki Brandta (1998), Starzyńskiej (2002) oraz tablice statystyczne Zielińskiego (1972) zawierają poprawne informacje związane z testem FS, a metodologia w nich zawarta jest zgodna z przedstawioną w niniejszej pracy. Mianowicie, jeśli numerujemy populacje tak, by $s_1^2 \geq s_2^2$ i liczymy wartość statystyki testowej zgodnie ze wzorem (2), to przy hipotezie alternatywnej $H_1 : \sigma_1^2 \neq \sigma_2^2$ przyjmujemy, przy poziomie istotności α, że H_1 jest prawdziwa, gdy $f \geq f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$, natomiast przy $H_1 : \sigma_1^2 > \sigma_2^2$, gdy $f \geq f_{1-a,\ n_1-1, n_2-1}$. Wątpliwości może budzić podręcznik Starzyńskiej (2002), w którym autorka w teście FS dopuszcza tylko jednostronną postać hipotezy alternatywnej. W tym przypadku możemy powiedzieć, że informacja o teście FS jest niepełna.

W podsumowaniu pracy autor (Drapella, 2016) zawarł dwa wnioski. Pierwszy jest sztuczny (w schemacie S_1 można pominąć losowość przydzielania wariancji do licznika i mianownika, pod którą kryje się numeracja prób, bowiem nie ma ona żadnego znaczenia), a drugi wniosek jest błędny.

Na zakończenie podano eksperyment numeryczny zbliżony do tego, który znajduje się w pracy Drapella, (2016) z poprawioną wartością granicy obszaru dopuszczalnego w schemacie S_2 oraz liczbą powtórzeń równą 10^6. Został on zaimplementowany w pakiecie statystycznym R (RCore Team, 2016) wspomaganym programem RStudio (RStudio Team, 2016). Z uwagi na jego prostotę, przedstawiono pełen kod:

```
set.seed(1313)
k1=0; k2=0 #zmienne pomocnicze
N=10^6 #liczba powtórzeń
n1=30; n2=30 #liczebności prób
alpha=0.1 #poziom istotności
q_d=qf(alpha/2,n1-1,n2-1) #kwantyl rzędu alpha/2
q_g=qf(1-alpha/2,n1-1,n2-1) #kwantyl rzędu 1-alpha/2
b=rbinom(N,1,1/2) #próba z rozkładu dwupunktowego 0,1
for (i in 1:N) {
x1=rnorm(n1); x2=rnorm(n2) #próby z rozkładu N(0,1)
f=var(x1)/var(x2) #wartość statystyki testowej z próby
########## schemat S1 ################################
if (b[i]==1) {f1=f} else {f1=1/f}
if (f1<=q_d | f1>=q_g) {k1=k1+1}
########## schemat S2 ################################
f2=max(c(f,1/f))
if (f2>=q_g) {k2=k2+1}}
k1/N #błąd I rodzaju w schemacie S1
## [1] 0.100053
k2/N #błąd I rodzaju w schemacie S2
## [1] 0.100053
```

Zgodnie z oczekiwaniami otrzymaliśmy w obu schematach taki sam błąd pierwszego rodzaju równy 0,100053, czyli bliski założonemu poziomowi istotności.

---

dr Wojciech Młocek — *Uniwersytet Rolniczy w Krakowie*

LITERATURA

Brandt, S. (1998). *Analiza danych*, PWN, Warszawa.

Drapella, A. (2016). O złej radzie dotyczącej testu F Snedecora, *Wiadomości Statystyczne*, nr 3, s. 61—67.

Fisher, R.A. (1925). *Statistical Methods For Research Workers*, Oliver and Boyd, Edinburgh.

RCore Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA URL http://www.rstudio.com.

Snedecor, G.W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance*, Collegiate Press, Ames, Iowa.

Starzyńska, W. (2002). *Statystyka praktyczna*, PWN, Warszawa.

Zieliński, R. (1972). *Tablice statystyczne*, PWN, Warszawa.

Summary. *In the study (Drapella, 2016), the author includes the thesis on erroneous method of calculating test statistics in the Fisher-Snedecor (FS) test. This method is commonly described in the literature, and examples, which cites the author, come from Brandt (1998) and Starzyńska (2002) textbooks as well as Zielinski's statistical tables (1972). The thesis is supported by numerical experiment and analytical considerations. This article indicates an error in the discussed study (Drapella, 2016), thus negates the included thesis, and shows that the said literature does not contain incorrect information associated with the FS test.*

Keywords: Fisher-Snedecor test.

Резюме. *В разработке (Драпелла, 2016) автор обсуждает тезис по неправильному способу расчета тестовых статистик в критерие Fishera--Snedecora (FS). Этот метод широко характеризуется в литературе, а примеры, которые представляет автор, происходят из учебников Брандта (1998), Стажиньской (2002) и статистических таблиц Зеліньского (1972). Тезис поддерживался численным экспериментом, а также аналитическими соображениями. В статье указано на ошибку в обсуждаемой разработке (Драпелла, 2016), таким образом, сводит на нет тезис в ней представленный и показывает, что в литературе нет неправильных информаций связанных с критерием FS.*

Ключевые слова: критерий Fishera-Snedecora.