

Testy proporcji w R. Test chi-kwadrat niezależności

Aleksander Zaigrajew



**UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU**

Wydział Matematyki
i Informatyki

Testy proporcji: jedna próbka

Testy proporcji można przeprowadzać dla jednej próbki i dla wielu próbek. Zaczynamy od przypadku jednej próbki.

Przykład 1. W losowo wybranej próbce 400 studentów znalazło się 144 palących. Interesuje nas frakcja p palących w populacji studentów.

1. Przeprowadź testowanie hipotezy $H_0: p = 0,3$ przeciw hipotezie $H_1: p > 0,3$, przyjmując poziom istotności testu $\alpha = 0,05$.
2. Zbuduj przedział ufności dla p na poziomie ufności 0,95.

Rozwiązanie. Funkcja **prop.test()** pozwala przeprowadzić test proporcji, jak też uzyskać przedział ufności dla tej proporcji. Ten test jest przybliżonym i opiera się na statystyce chi-kwadrat.

Ważną opcją jest używanie (bądź nie) **correct=T**. Za pomocą **correct=T** (domyślnie) uzyskujemy tzw. *poprawkę na ciągłość*. Taka poprawka się stosuje, gdy przynajmniej jedna z liczebności oczekiwanych jest bardzo mała (można stosować już przy liczebnościach mniejszych niż 10, a na pewno przy mniejszych niż 5).

Testy proporcji: jedna próbka (2)

W naszym przypadku nie na potrzeby stosować poprawkę na ciągłość, bowiem liczebności oczekiwane przy prawdopodobieństwie $p = 0,3$ w próbie rozmiaru $n = 400$ wynoszą:

$$p \cdot n = 0,3 \cdot 400 = 120 \text{ oraz } (1 - p) \cdot n = (1 - 0,3) \cdot 400 = 280.$$

Zwróćmy uwagę, że hipoteza alternatywna jest jednostronna (prawostronna). Zatem wynik testowania hipotez uzyskujemy za pomocą komendy:

```
> prop.test(144, 400, 0.3, alternative = "g", correct = F)
```

Skoro p -wartość jest równa 0.004414, odrzucamy hipotezę zerową.

Przedział ufności dla p uzyskujemy tak:

```
> prop.test(144, 400, correct = F)
```

Wynik: $[0.3144979, 0.4081655]$.

Testy proporcji: jedna próbka (3)

Zauważmy, że gdy hipoteza alternatywna jest dwustronna, czyli postaci $H_1: p \neq 0,3$, można użyć innego sposobu, mianowicie testu chi-kwadrat:

```
> x = c(rep(0, 256), rep(1, 144))  
> chisq.test(table(x), p = c(0.7, 0.3))
```

który daje dokładnie ten sam wynik, co

```
> prop.test(144, 400, 0.3, correct = F)
```

z p -wartością równą 0.008829 (pierwsze polecenie tworzy próbę x rozmiaru 400 z 256 zerami oraz 144 jedynkami).

Inny sposób, jeszcze bardziej przybliżony, to używanie testu t -Studenta do testowania proporcji:

```
> t.test(x, mu = 0.3, alternative = "g")
```

P -wartość jest równa 0.006466, więc odrzucamy hipotezę zerową.

Testy proporcji: jedna próbka (4)

Przykład 2. Wśród 300 ankietowanych osób odpowiedź „tak” padła w 30 przypadkach. Czy można powiedzieć, że wśród ogółu populacji tylko 7% tak odpowiada? Jaki jest przedział ufności dla prawdopodobieństwa uzyskania odpowiedzi „tak” od losowo wybranej osoby z populacji?

Rozwiązanie. Testujemy hipotezy: H_0 : 7% populacji odpowiada „tak” (czyli $p = 0,07$), H_1 : inny procent populacji odpowiada „tak” ($p \neq 0,07$). Zatem

```
> prop.test(30, 300, 0.07, correct = F)
```

P -wartość wynosi 0.0417, więc na poziomie istotności $\alpha = 0.05$ odrzucamy H_0 . Przedział ufności dla p : [0.07094791, 0.13916646].

Uwaga. Gdyby stosowalibyśmy test proporcji z poprawką na ciągłość bądź test t -Studenta wynik testowania byłby inny!

Testy proporcji: wiele próbek

Test proporcji da się stosować nie tylko wtedy, gdy mamy jedną próbkę, ale też wtedy, gdy próbek jest więcej.

Przykład 3. Niech wśród 300 ankietowanych osób odpowiedź „tak” padła w 30 przypadkach. W innej 200-osobowej grupie odpowiedź „tak” padła w 25 przypadkach, w trzeciej 500-osobowej grupie odpowiedź „tak” padła w 40 przypadkach. Czy te 3 próby pochodzą z tej samej populacji?

Rozwiązanie. Rozważamy hipotezy: H_0 : próby pochodzą z tej samej populacji (czyli $p_1 = p_2 = p_3$), H_1 : próby nie pochodzą z tej samej populacji .

(Tutaj p_i to proporcja odpowiadających „tak” w i -tej populacji; jeśli $p_1 = p_2 = p_3$, to mamy do czynienia z jedną dużą populacją).
Zatem:

$> \text{prop.test}(c(30, 25, 40), c(300, 200, 500))$

Skoro p -wartość wynosi 0.1747, na poziomie istotności $\alpha = 0.05$ nie ma podstaw do odrzucenia H_0 .

Test niezależności: przykłady

Ostatni test, który rozważymy, nazywa się **testem niezależności chi-kwadrat** i testuje niezależność/zależność dwóch cech.

Przykład 4. Rzecznik pewnej partii (oznaczymy ją A) twierdzi, że wśród zwolenników tej partii, miłośnicy muzyki disco-polo, rockowej i symfonicznej występują mniej więcej w tych samych proporcjach, co w całej populacji wyborców (co by mówiło o niezależności dwóch cech wyborcy: *stosunku do partii A i preferencji muzycznych*). Przeprowadzono sondaż. Wśród wylosowanych 100 wyborców wyniki badania były następujące:

	Popieram A	Nie popieram A	Razem
Słucham disco-polo	25	10	35
Słucham muzyki rockowej	20	20	40
Słucham muzyki symfonicznej	15	10	25
Razem	60	40	100

Test niezależności: przykłady (2)

Rozwiązanie. Rozważamy następujące hipotezy:

H_0 : cechy są niezależne, H_1 : cechy są zależne.

```
> chisq.test(data.frame(c(25, 20, 15), c(10, 20, 10)))
```

P-wartość wynosi 0.1677, zatem na poziomie istotności $\alpha = 0.05$ nie ma podstaw do odrzucenia H_0 (czyli rzecznik ma rację).

Przykład 5. Niech w grupie pacjentów, przyjmujących nowy lek, 19 pozostało bez poprawy, 41 odnotowało wyraźną poprawę, 60 osób całkowicie wyzdrowiało. W grupie kontrolnej leczonej dotychczasowymi lekami wartości te wynosiły odpowiednio 46, 19, 15.

Czy nowy lek faktycznie jest lepszy?

Rozwiązanie. Rozważamy hipotezy: H_0 : leki dają ten sam efekt (czyli wyniki nie zależą od tego, czy stosujemy nowy lek czy stary), H_1 : leki dają różny efekt (czyli wyniki zależą od tego, jaki lek stosujemy; patrząc na dane oznacza to, że nowy lek jest lepszy).

```
> chisq.test(data.frame(c(19, 41, 60), c(46, 19, 15)))
```

P-wartość ≈ 0 , więc na poziomie istotności 0.05 odrzucamy H_0 .

- Wojciech Niemirow, **Rachunek prawdopodobieństwa i statystyka matematyczna**, SNS, Warszawa, 1999
- Przemysław Biecek, **Przewodnik po pakiecie R**, Oficyna Wydawnicza GiS, Wrocław, 2011
- Łukasz Komsta, **Wprowadzenie do środowiska R**