Porównanie funkcji mocy testów statystycznych w różnych warunkach

Amelia Bieda Karolina Bakalarz

Wprowadzenie

W analizie statystycznej często stajemy przed problemem testowania hipotez dotyczących wartości oczekiwanej populacji. W tym celu wykorzystuje się różne testy statystyczne, w zależności od założeń dotyczących rozkładu danych i dostępnych informacji o populacji.

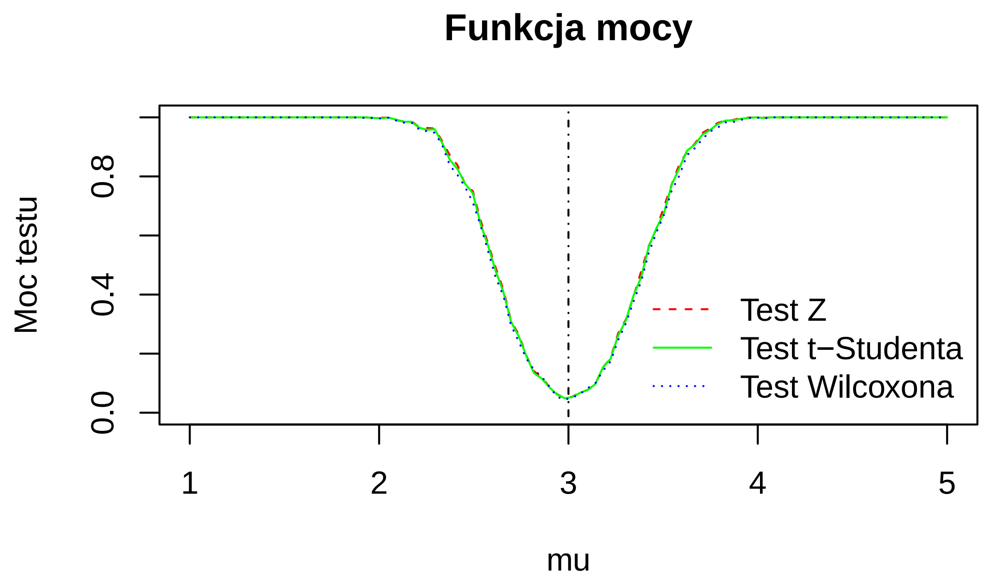
W niniejszej analizie rozważamy testy na poziomie istotności $\alpha = 0.05$. Zastosowano trzy testy do weryfikacji hipotezy zerowej $H_0 : \mu = 3$, przeciwko hipotezie alternatywnej $H_1 : \mu \neq 3$.

1. Test Z przy założeniu $\sigma = 2$ - wykorzystywany przy znanej wariancji populacji, bazuje na rozkładzie normalnym statystyki testowej;
2. Test Studenta - stosowany przy nieznanej wariancji, którą należy oszacować na podstawie próby, sprawdza czy średnia populacji istotnie się różni od wartości z hipotezy.
3. Test rang znakowanych Wilcoxona - jest testem nieparametrycznym, z którego korzystamy, gdy nie można założyć normalności rozkładu badanych danych, dzięki porównywaniu median, a nie średnich jest odporniejszy na odchylenia od założeń normalności.

Porównanie funkcji mocy tych testów pozwala ocenić ich skuteczność w wykrywaniu różnic względem wartości oczekiwanej w populacji oraz określić, czy istnieje test jednostajnie najmocniejszy w danym przypadku. Dzięki implementacji testów w języku R możemy za pomocą symulacji Monte Carlo przeanalizować funkcje mocy oraz prawdziwość postawionych hipotez.

Zadanie 1

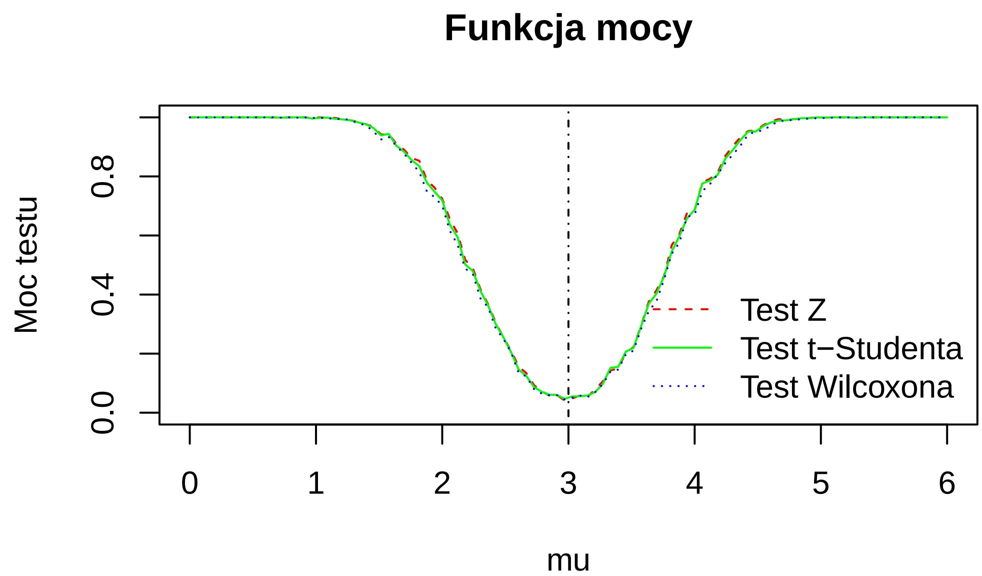
Rozważmy próbę $(X_1, ..., X_{100})$ z rozkładu normalnego $N(\mu, 2^2)$. Korzystając z symulacji Monte Carlo wykonaj wykres funkcji mocy w zależności od μ na przedziale $(1, 5)$ dla wszystkich trzech testów. Czy istnieje test jednostajnie najmocniejszy spośród nich?



W badanym przypadku nie istnieje test jednostajnie najmocniejszy z wykorzystanych testów. Wszystkie trzy są bardzo skuteczne - przy wartości μ bliskiej 3 moce testów są niskie, co oznacza rzadkie, prawdopodobnie bliskie zeru szanse na odrzucenie hipotezy zerowej. Wraz z oddalaniem się wartości μ od zadanego w hipotezie zerowej $\mu = 3$, rośnie funkcja mocy dla wszystkich wykorzystanych testów, co oczywiście oznacza zwiększenie się prawdopodobieństwa odrzucenia hipotezy zerowej na korzyść hipotezy alternatywnej $\mu \neq 3$.

Zadanie 2

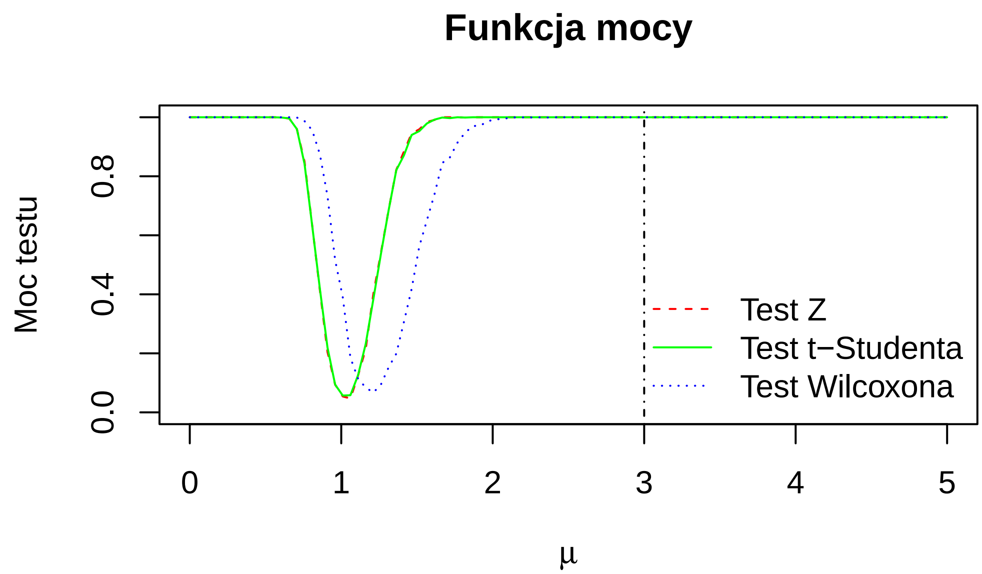
Rozważmy próbę $(X_1, ..., X_{100})$ z rozkładu normalnego $N(\mu, 4^2)$. Wykonaj wykres funkcji mocy na wybranym przedziale zawierającym przynajmniej po jednym punkcie z hipotezy zerowej i alternatywnej. Jak zmieniła się funkcja mocy testów? Czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród nich?



Tak jak w poprzednim przypadku, moc testów jest najniższa w okolicach trójki co oznacza, że rzadko odrzucają one hipotezę zerową, gdy jest ona prawdziwa. Im dalej od trójki tym większa moc testów, szczególnie dla $\mu < 1$ oraz $\mu > 5$ prawdopodobieństwo przyjęcia hipotezy alternatywnej jest bliskie lub nawet równe 1. Zauważalny jest wpływ zwiększonej wariancji w porównaniu do poprzedniego zadania - funkcja mocy rośnie wolniej. Dla wartości μ bliskich 3 testy mają niższą moc, zatem trudniej jest wykryć niewielkie odchylenia od hipotezy zerowej. Nie istnieje test jednostajnie najmocniejszy w rozważanym zadaniu.

Zadanie 3

Rozważmy próbę $(X_1, ..., X_{100})$ z rozkładu wykładniczego $Exp(1/\mu)$. Wykonaj wykres funkcji mocy na wybranym przedziale zawierającym przynajmniej po jednym punkcie z hipotezy zerowej i alternatywnej. Jak zmieniła się funkcja mocy testów? Czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród nich?



Wykres funkcji mocy testu Z i t-Studenta są identyczne, z najmniejszą wartością blisko jedynki. Średnia rozkładu wykładniczego o parametrze skali $1/\mu$ wynosi μ, jednak testy są przystosowane dla rozkładu normalnego, stąd dla asymetrycznego rozkładu wykładniczego nie są odporne i wychodzi większe prawdopodobieństwo popełnienia błędu I i II rodzaju. Wariancja tego rozkładu wynosi $1/\mu^2$, więc dla mniejszych wartości μ wariancja jest większa, co również utrudnia wykrycie różnic między hipotezami. Test Wilcoxona rośnie wolniej niż testy parametryczne, ma najmniejszą wartość przy 1.2 (bliżej trójki niż poprzednie testy). Jest mniej wrażliwy na skośność rozkładu i nie zakłada normalności. Mimo nienakładania się funkcji mocy testów parametrycznych z testem nieparametrycznym, nie istnieje w tym przypadku test jednoznacznie najmocniejszy.

Podsumowanie

Dla normalnych danych i dużych próbek wszystkie trzy testy dają bardzo podobne wyniki. Większa wariancja zmniejsza moc testów parametrycznych przez mniejszą zdolność do wykrywania małych odchyleń od hipotezy zerowej. Test Wilcoxona działa dobrze dla danych normanych, a dla danych z rozkładu wykłaniczego jest dokładniejszy od testów parametrycznych, jednak nie jest idealny i jego moc również zależy od struktury danych. Przy niespełnionych założeniach testów dużo łatwiej o błędne wnioski dla H_0 i H_1.