Inżynierskie zastosowania statystyki - Ćwiczenia nr 6

1 Wstęp teoretyczny

1.1 Testy istotności dla wartości średniej (c.d.)

W testach do weryfikacji hipotezy o wartości przeciętnej $H_0 : \mu = \mu_0$ wykorzystujemy statystykę $\bar{X}$, czyli średnią arytmetyczną próby. Jeżeli wariancja rozkładu cechy w populacji generalnej σ^2 nie jest znana, zastępujemy ją wartością estymatora wariancji

$$s^2 = \frac{1}{n-1} \sum_{i=1} n x_i^2. \tag{1}$$

1.1.1 Model 2.

Badana cecha X populacji ma rozkład $N(\mu, \sigma^2)$ przy nieznanej wariancji σ^2.

Statystyka testowa

$$t = \frac{\bar{X} - \mu_0}{s} \sqrt{n}, \tag{2}$$

przy założeniu prawdziwości hipotezy $H_0 : \mu = \mu_0$, ma rozkład Studenta o $v = n - 1$ stopniach swobody ($t_v = t_{n-1}$).

1.1.2 Model 3.

Badana cecha X populacji ma rozkład dowolny o nieznanej wartości średniej μ i nieznanej wariancji σ^2 oraz n jest duże ($n \geq 100$).

Statystyka testowa

$$Z = \frac{\bar{X} - \mu_0}{s} \sqrt{n}, \tag{3}$$

przy założeniu prawdziwości hipotezy $H_0 : \mu = \mu_0$, ma asymptotyczny rozkład normalny $N(0, 1)$.

1.2 Test znaków

Badana cecha X populacji ma dowolny rozkład ciągły o nieznanej medianie m_e. Niech hipotetyczna wartość mediany wynosi m_0 i testowana jest hipoteza zerowa $H_0 : m_e = m_0$.

Statystyka testowa R^+ oznacza ilość obserwacji dla których $x_i - m_0 > 0$.

1.3 Test znakowanych rang Wilcoxona

Badana cecha X populacji ma rozkład ciągły i symetryczny o nieznanej medianie $m_e = \mu$. Statystyka testowa $w = \min(w^+, w^-)$, w^+ oznacza sumę rang w_i dla których $x_i - \mu_0 > 0$ a w^- oznacza sumę wartości bezwzględnych rang w_i dla których $x_i - \mu_0 < 0$.

1.4 p-wartość

Definicja 1 *p-wartość (p-value) - prawdopodobieństwo, że przy prawdziwości hipotezy zerowej statystyka testowa przyjmie wartość bardziej ekstremalną niż ta, która została zaobserwowana.*

1.5 Ogólna procedura testowania (wersja II)

1. Wskaż parametr, którego test ma dotyczyć.
2. Postaw hipotezę zerową H_0.
3. Postaw hipotezę alternatywną H_a (H_1).
4. Ustal poziom istotności α.
5. Wskaż statystykę testową.
6. Oblicz wartość statystyki.
7. Oblicz p-wartość.
8. Zdecyduj czy hipoteza zerowa powinna zostać odrzucona i skomentuj to w kontekście problemu.

2 Zadania

1. Korzystając z wykresów na rysunku 1 porównaj rozkład statystyki t ze standaryzowanym rozkładem normalnym. Uwzględnij wartość średnią rozkładu, wariancję rozkładu, symetrię. Czy wielkość próby ma znaczenie dla rozkładu statystyki t?

2. Kwantyle rozkładu Studenta oznaczane są przez $t_{p,v}$ gdzie p oznacza rząd kwantyla a v liczbę stopni swobody rozkładu. Oczytaj z tablicy rozkładu Studenta wartości kwantyli $t_{0.025,15}$, $t_{0.05,10}$, $t_{0.1,20}$, $t_{0.005,25}$.

3. Testowana jest hipoteza $H_0 : \mu = 5$ przeciwko

   (a) $H_1 : \mu \neq 5$
   (b) $H_1 : \mu > 5$
   (c) $H_1 : \mu < 5$

   na podstawie 20-elementowej próby. Niech $\bar{X} = 4.5$, $s^2 = 1.44$. Odczytaj p-wartość dla każdej z trzech par hipotez.

4. W celu oszacowania zanieczyszczenia rtęcią u karpi pobrano próbę 144 karpi ze stawów hodowlanych na Dolnym Śląsku i zmierzono zawartość rtęci w tkance mięśniowej każdej z ryb. Średnia arytmetyczna zawartości rtęci wynosi $\bar{x} = 2.3635$, a odchylenie standardowe z próby $s = 0.7076$. Zgodnie z normą stężenie rtęci u karpia nie powinno przekraczać $2.5\frac{\mu g}{100g}$. Czy na poziomie istotności $\alpha = 0.05$ średnie stężenie rtęci u dolnośląskich karpi nie przekracza normy? Oblicz p-wartość dla tego testu.

5. Poddano badaniu zawartość sodu w $n = 30$ 300-gramowych paczkach płatków kukurydzianych. Na podstawie danych (w mg) obliczono średnią z próby $\bar{x} = 129.7527$ oraz wariancję z próby $s^2 = 0.86371$. Czy można twierdzić, że paczka płatków kukurydzianych tej marki zawiera średnio 130 mg sodu? Oblicz p-wartość dla tego testu.

6. Zmierzono wytrzymałość 10 losowo wybranych gotowych elementów konstrukcji budowlanych i otrzymano następujące wyniki (w MPa): 383, 284, 339, 340, 305, 386, 378, 335, 344, 346. Zakładając, że rozkład wytrzymałości tych elementów jest rozkładem $N(\mu, \sigma^2)$ o nieznanych parametrach, przetestuj hipotezę zerową $H_0 : \mu = 340$ przeciwko hipotezie alternatywnej $H_1 : \mu < 340$ na poziomie istotności 0.05. Oblicz p-wartość dla tego testu.

7. Pobrano i zmierzono wartość pH dziesięciu próbek kąpieli galwanicznej używanej w procesie produkcji elektroniki. Wartości pH próbek wyniosły 7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, 7.42. Technik produkcji uważa, że pH ma medianę równą 7.0. Czy dane uzyskane z próby wskazują, że to stwierdzenie jest prawdziwe? Zweryfikuj hipotezę za pomocą testu znaków. Określ p-wartość dla tego testu.

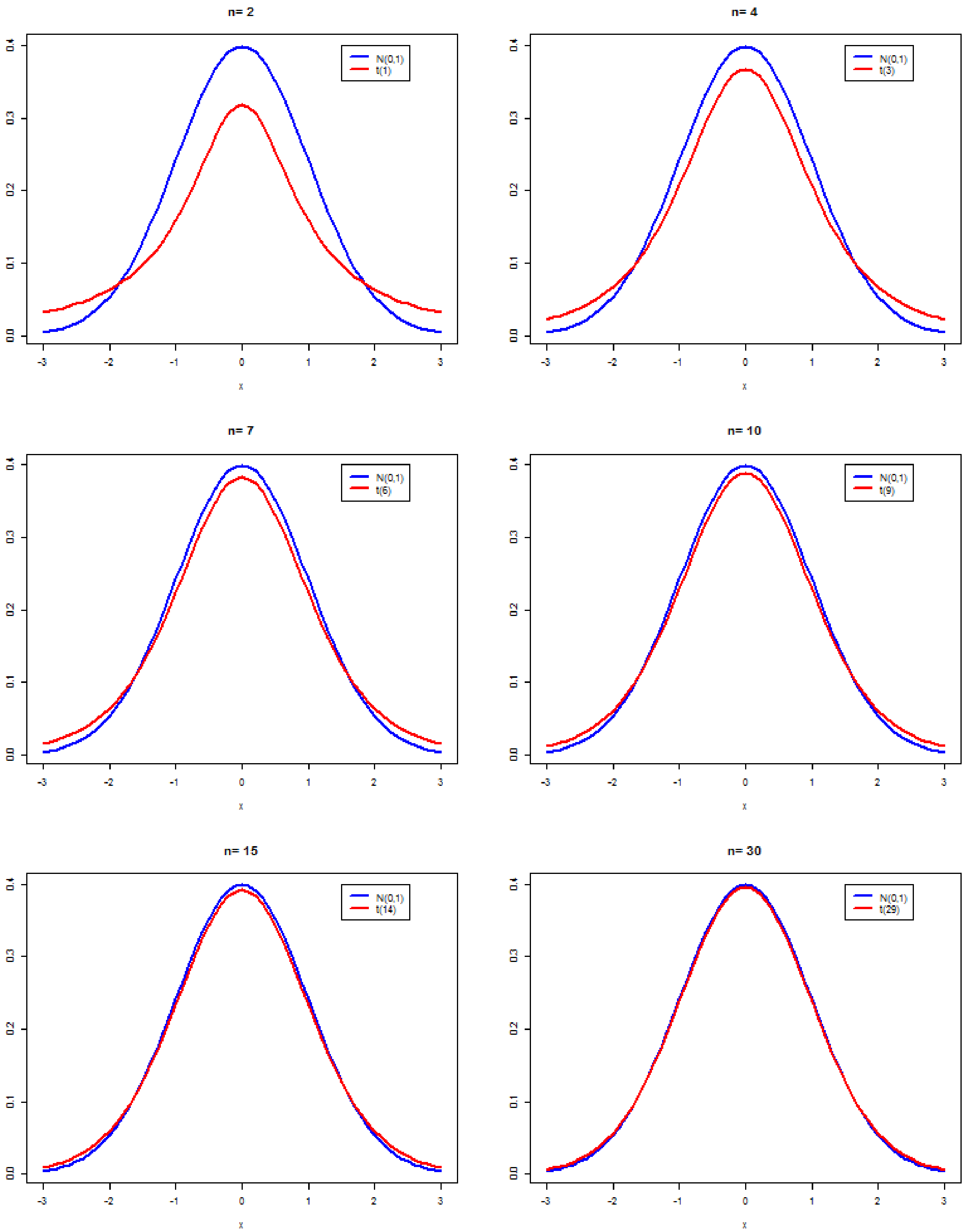

Figure 1: Rozkład t-Studenta (zad 1)

8. Zawartość tytanu w stopach aluminium używanych w przemyśle lotniczym jest ważnym wyznacznikiem siły. Próba 20 kuponów blachy alumionowej wykazała następującą zawartość tytanu (w procentach): 8.32, 8.05, 8.93, 8.65, 8.25, 8.46, 8.52, 8.35, 8.36, 8.41, 8.42, 8.30, 8.71, 8.75, 8.60, 8.83, 8.50, 8.38, 8.29, 8.46. Mediana zawartości tytanu powinna być równa 8.5%. Postaw odpowiednie hipotezy i zweryfikuj je za pomocą testu znaków. Określ p-wartość dla testu.

9. Rozważ dane z zadania 7. Załóż, że rozkład pH jest symetryczny i ciągły. Za pomocą testu znakowanych rang Wilcoxona zweryfikuj hipotezę $H_0 : \mu = 7$ przeciwko $H_1 : \mu \neq 7$ na poziomie istotności $\alpha = 0.05$.

10. Rozważ dane z zadania 8. Załóż, że rozkład zawartości tytanu jest symetryczny i ciągły. Za pomocą testu znakowanych rang Wilcoxona zweryfikuj hipotezę $H_0 : \mu = 8.5$ przeciwko $H_1 : \mu \neq 8.5$ na poziomie istotności $\alpha = 0.05$.