Przedziały ufności. Rozkład t. Test t studenta

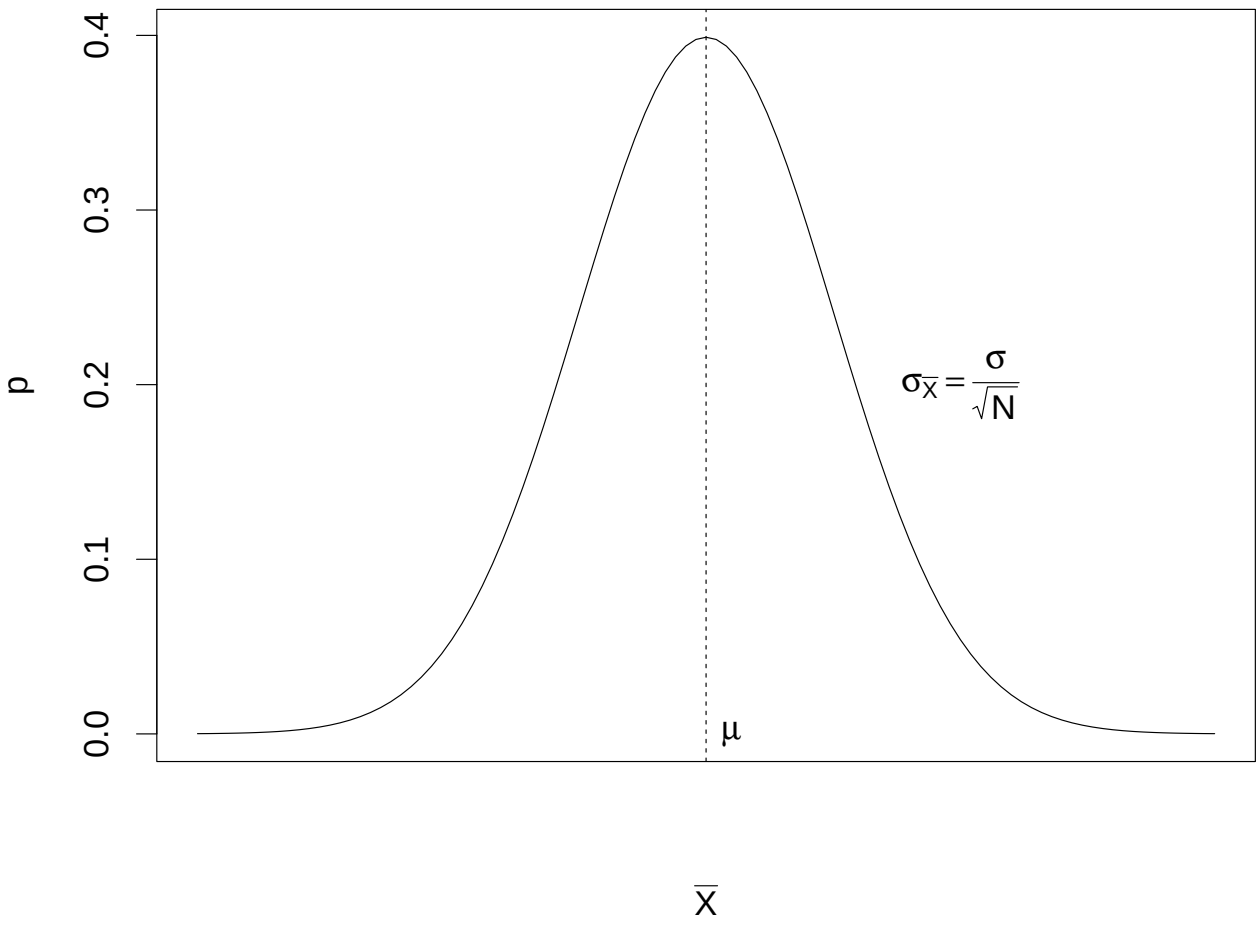
Centralne Twierdzenie Graniczne - przypomnienie

Proste symulacje w R pokażą nam to, co teoretycznie wiemy z Centralnego Twierdzenia Granicznego:

- że średnie z prób losowych oscylują wokół średniej w populacji,
- a w szczególności, że niezależnie od rozkładu zmiennej w populacji średnie z prób losowych rozkładają się normalnie wokół tej średniej,
- że ich rozproszenie jest tym mniejsze, im mniejsze jest rozproszenie zmiennej w populacji,
- i tym mniejsze, im większa jest liczebność prób,
- a dokładniej, że odchylenie standardowe średnich z prób losowych równe jest odchyleniu standardowemu w populacji podzielonemu przez pierwiastek liczebności prób.

Taki rozkład teoretyczny $N(\mu, \sigma^2/N)$, który nazywamy rozkładem średniej z próby, jest rozkładem prawdopodobieństwa, który pozwala nam szacować, jak prawdopodobne jest wylosowanie z danej populacji próby o takiej czy innej średniej. Odchylenie standardowe tego rozkładu nazywa się błędem standardowym średniej.

```
par(cex = 1.8)
curve(dnorm(x), from = -4, to= 4,
     main = "Rozkład średniej z próby",
     ylab = 'p',
     xlab = expression(bar(X)),
     xaxt = 'n')
abline(v = 0, lty = 2)
text(x=0+0.2, y = 0,
     labels = expression(mu))
text(x= 2, y = 0.2,
     labels = expression(sigma[bar(X)] == frac(sigma, sqrt(N))))
```




```
# Polecenie `expression` pozwala nam umieszczać na wykresach wyrażenia matematyczne.
# Składnia używana przez R jest dość podoba do LaTeXa.
# Dokumentacja znajduje się tutaj:
# https://stat.ethz.ch/R-manual/R-devel/library/grDevices/html/plotmath.html
```

Pytanie

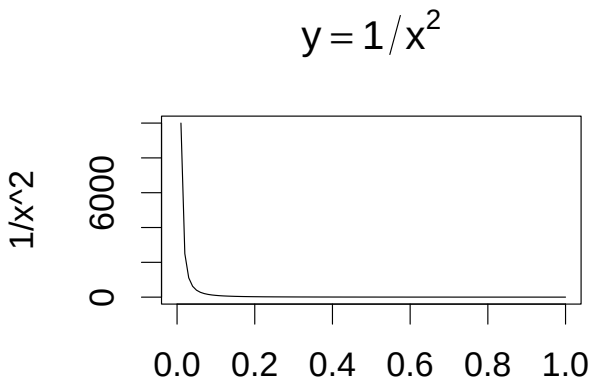
Oczywiście zależy nam na tym, żeby błąd standardowy był jak najmniejszy: dlatego dążymy do dużego N. Z drugiej strony ten zysk na precyzji jest szczególnie duży przy małych wartościach N; im większe N, tym mniejsza redukcja błędu standardowego (to też możemy zobrazować za pomocą prostej symulacji).

Dlaczego? Która z poniższych funkcji oddaje zależność między N a błędem standardowym średniej?

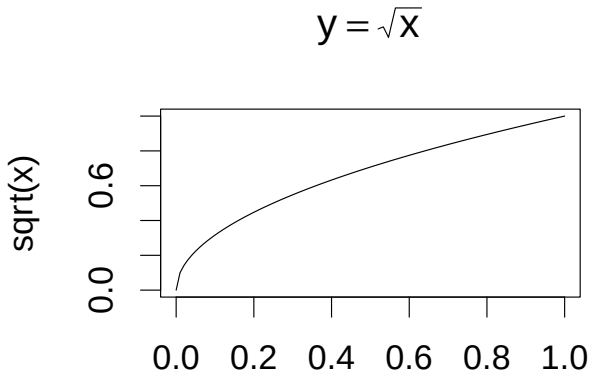
Odpowiedź

Dlatego że w mianowniku jest pierwiastek. Wykres numer 3.

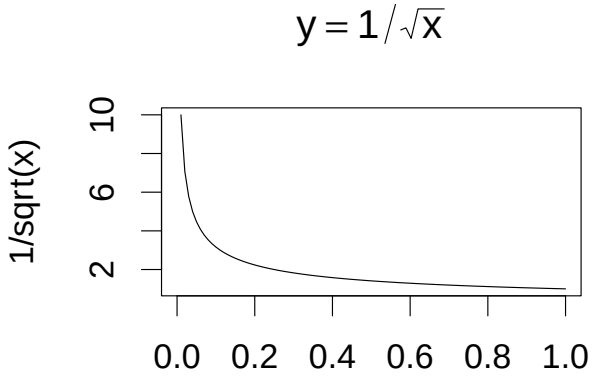
```
par(mfrow = c(2,2))
par(cex = 1.8)
curve(1/x**2, main = expression(y == 1/x^2), sub = "Wykres 1")
curve(sqrt(x), main = expression(y== sqrt(x)), sub = "Wykres 2")
curve(1/sqrt(x), main = expression(y == 1/sqrt(x)), sub = "Wykres 3")
curve(x**2, main = expression(y==x^2), sub = "Wykrest 4")
```



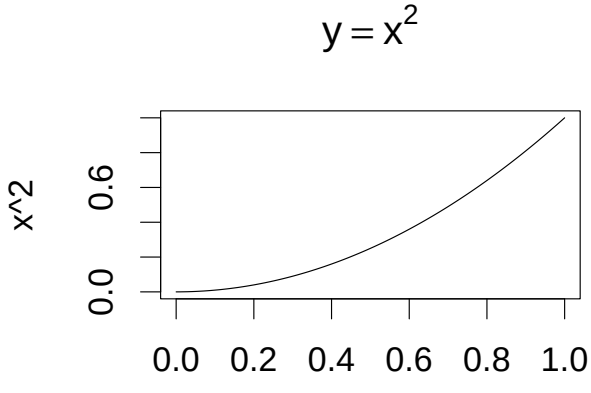
Wykres 1



Wykres 2



Wykres 3

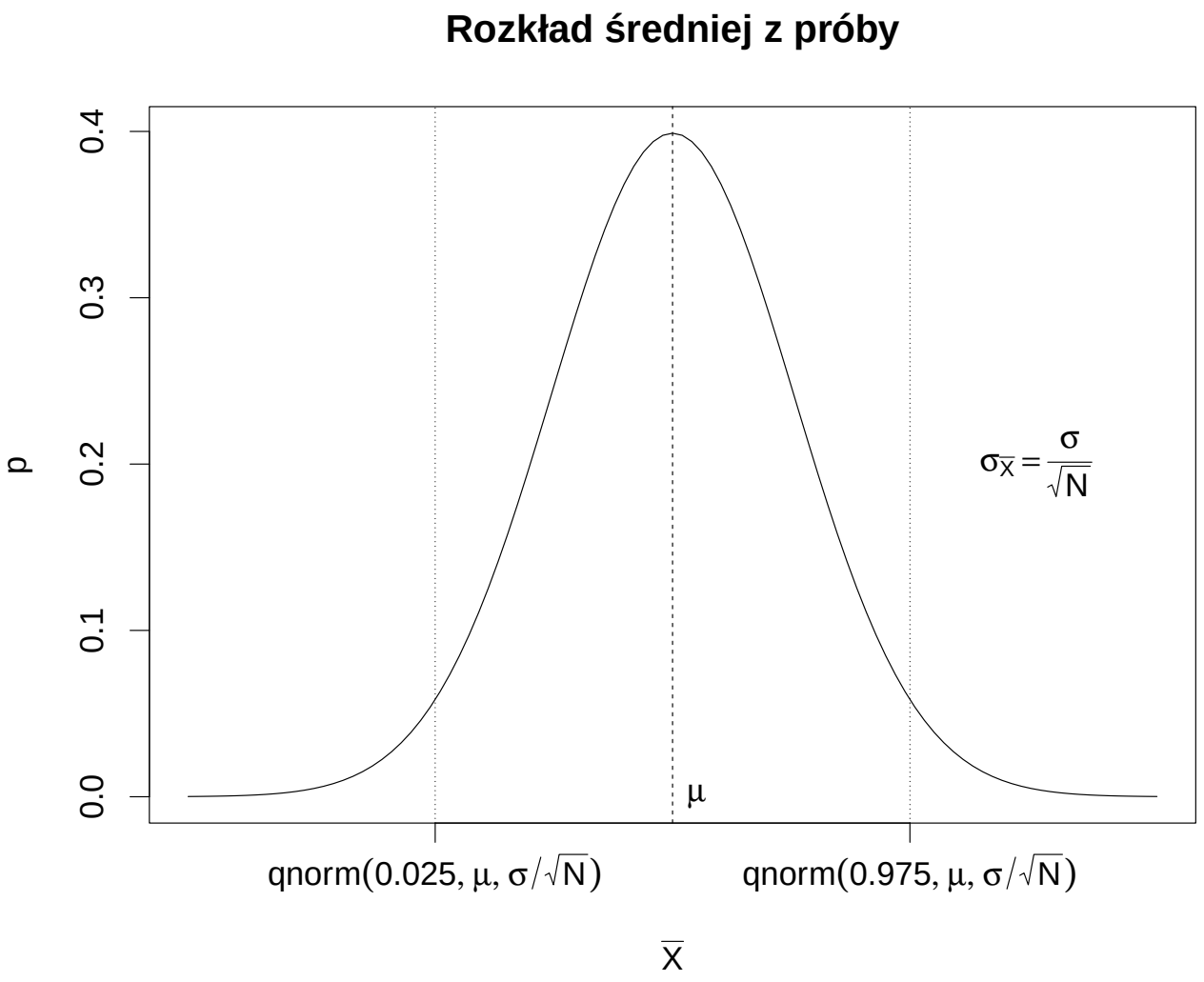


Wykrest 4

Potrafiąc skonstruować rozkład średniej z próby, możemy testować prostą hipotezę zerową: dana próba losowa pochodzi z populacji o takiej i takiej średniej. Jeśli różnica między średnią naszej próby a założoną zgodnie z hipotezą zerową średnią w populacji jest zbyt duża (jeśli nasza średnia wpada w obszar krytyczny), odrzucamy hipotezę zerową.

```
par(cex = 1.8)

curve(dnorm(x), from = -4, to= 4,
     main = "Rozkład średniej z próby",
     ylab = 'p',
     xlab = expression(bar(X)),
     xaxt = 'n')
abline(v = 0, lty = 2)
abline(v=qnorm(0.025), lty = 3)
abline(v=qnorm(0.975), lty = 3)

text(x=0+0.2, y = 0,
     labels = expression(mu))
text(x= 3, y = 0.2,
     labels = expression(sigma[bar(X)] == frac(sigma, sqrt(N))))
```

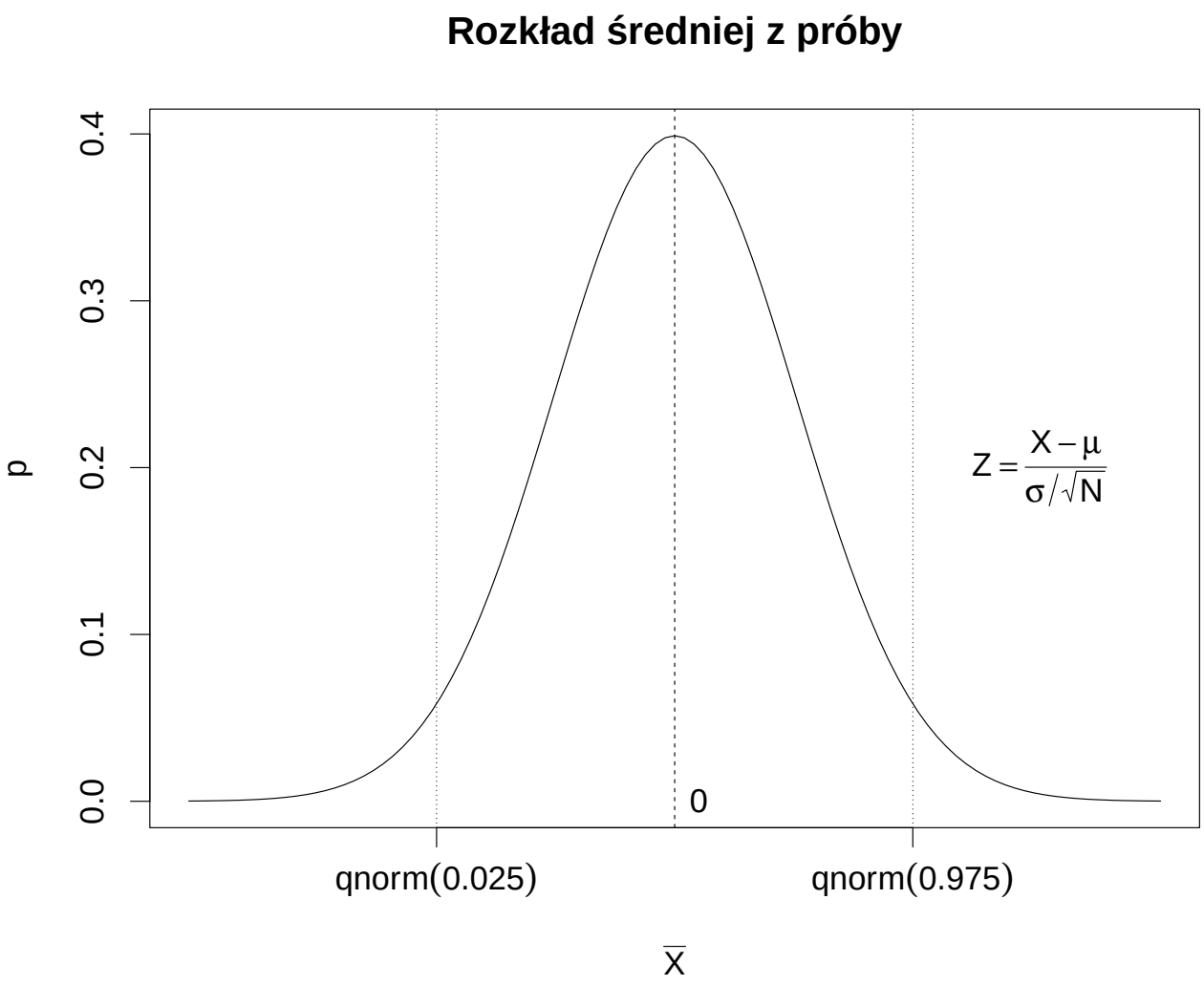
```
lab1 <- expression(qnorm(0.025, mu, sigma/sqrt(N)))
lab2 <- expression(qnorm(0.975, mu, sigma/sqrt(N)))

axis(side =1 , at=qnorm(c(0.025, 0.975)), labels= c(lab1, lab2))
```



Wygodniej może być wystandaryzować rozkład średniej z próby, czyli doprowadzić go do takiej postaci, w której średnia wynosi 0, a odchylenie standardowe — 1. Taki rozkład "działa" zawsze tak samo, niezależnie od oryginalnej skali naszej zmiennej.

```
par(cex = 1.8)

curve(dnorm(x), from = -4, to= 4,
     main = "Rozkład średniej z próby",
     ylab = 'p',
     xlab = expression(bar(X)),
     xaxt = 'n')
abline(v = 0, lty = 2)
abline(v=qnorm(0.025), lty = 3)
abline(v=qnorm(0.975), lty = 3)

text(x=0+0.2, y = 0, labels = expression(0))
```

```
text(x= 3, y = 0.2, labels = expression(Z == frac(X -mu,sigma/sqrt(N))))


lab1 <- expression(qnorm(0.025))
lab2 <- expression(qnorm(0.975))

axis(side =1 , at=qnorm(c(0.025, 0.975)), labels= c(lab1, lab2))
```



Pytanie

Zakładając, że średnia naszej próby to `srednia`, jej liczebność to `N`, odchylenie standardowe w populacji to `sigma`, a przyjęta zgodnie z hipotezą zerową średnia w populacji to `mu0`, jakiego wyrażenia możemy użyć, żeby sprawdzić, czy nasza średnia wpada w pięcioprocentowy obszar krytyczny?

- `srednia < qnorm(.025, mu0, sigma/sqrt(N)) | qnorm(.975, mu0, sigma/sqrt(N)) < srednia`
- `srednia < mu0 - qnorm(.975) * sigma/sqrt(N) | mu0 + qnorm(.975) * sigma/sqrt(N) < srednia`
- `srednia < mu0 + qnorm(.025) * sigma/sqrt(N) | mu0 - qnorm(.025) * sigma/sqrt(N) < srednia`
- `srednia < mu0 - qnorm(.025) * sigma/sqrt(N) | mu0 + qnorm(.975) * sigma/sqrt(N) < srednia`

Odpowiedź

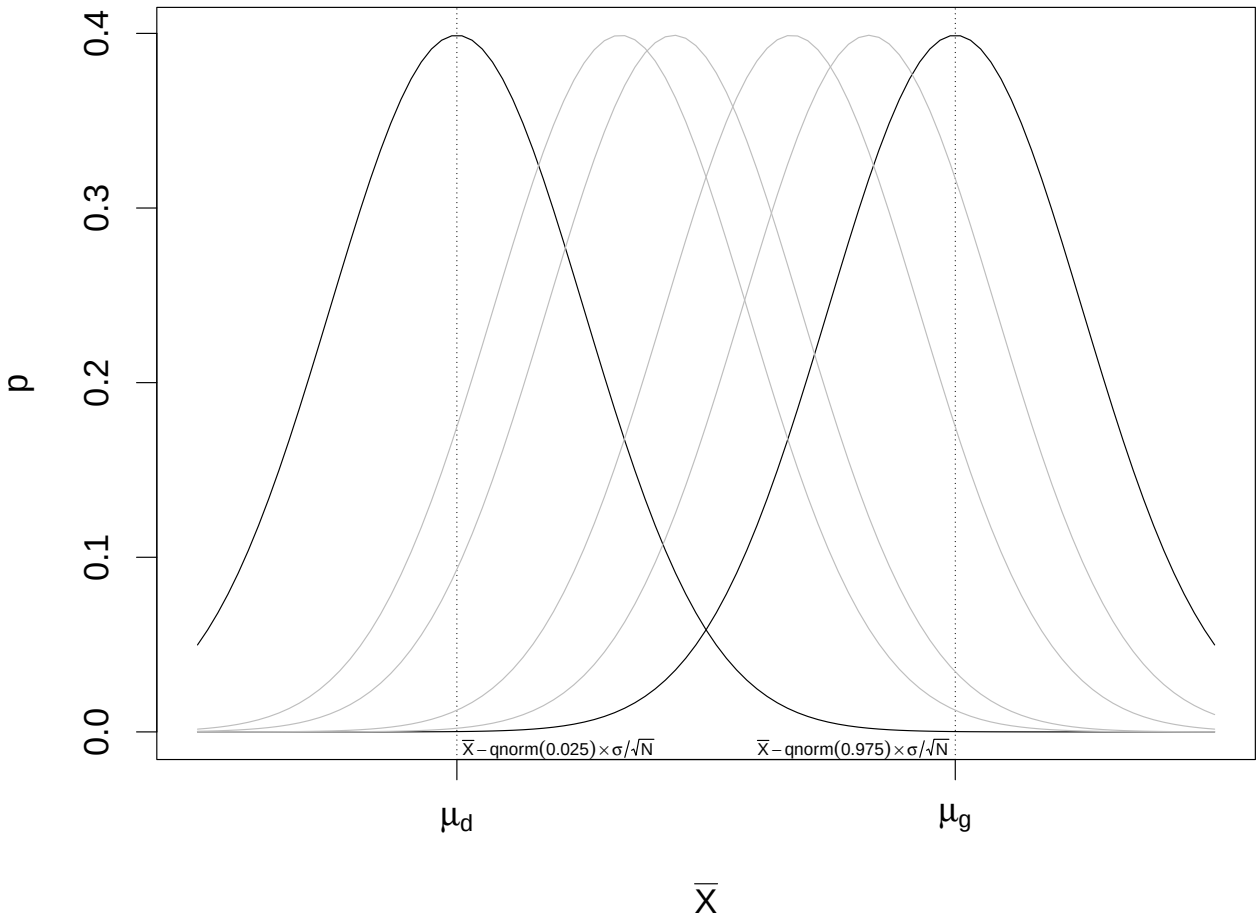
Wszystkie odpowiedzi są prawidłowe, oprócz ostatniej. W pierwszej po prostu ustawiamy parametry rozkładu. W drugiej i trzeciej korzystamy z faktu, że możemy przekształcic dowolny rozkład normalny w standardowy rozkład normalny i odwrotnie. Trzecia jest w zasadzie identyczna z druga, tylko odwracamy znak. W czwartej otrzymujemy dwie takie same wartości - alternatywa jest więc bez sensu.

Przedziały ufności

Alternatywą dla testowania hipotez (albo ich dopełnieniem; zależy, jak na to patrzeć) jest konstruowanie przedziałów ufności wokół statystyki z próby. Podręczniki i kursy, a w konsekwencji badacze w praktyce, skupiają się na tym pierwszym, a niesłusznie, bowiem przedziały ufności dają nam więcej informacji. Testowanie hipotezy zerowej pomaga jedynie odpowiedzieć na pytanie o jakąś pojedyńczą hipotetyczną wartość średniej w populacji. Skonstruowanie przedziału ufności pozwala określić cały zakres wartości (μD, μG), który z określonym prawdopodobieństwem obejmuje średnią populacji.

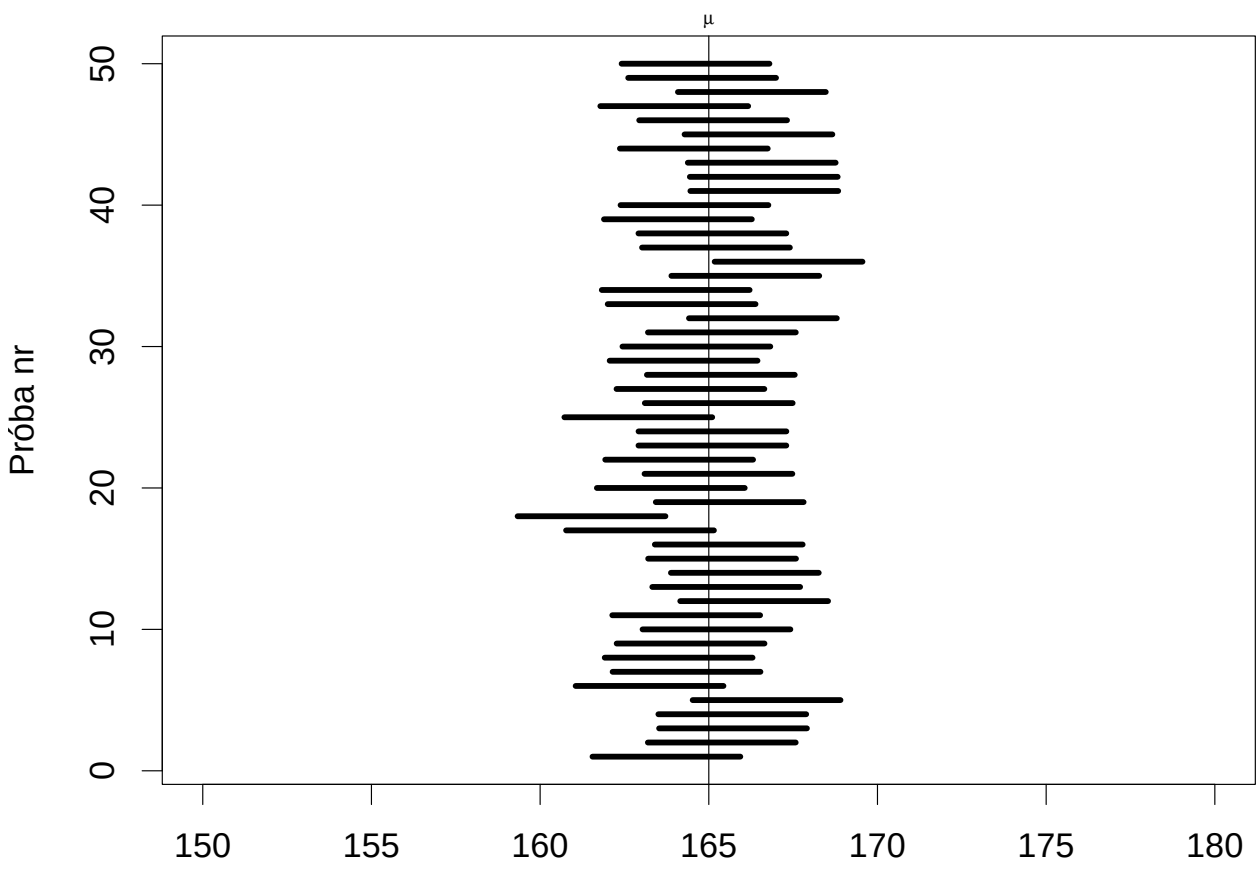
Poniżej znajduje się wykres na którym przedstawione mamy rozkłady średniej z próby dla różnych hipotez zerowych. Nie odrzucamy hipotezy zerowej dla wszystkich tych hipotez, dla których średnia populacji mieści się miedzy (μD, μG).

```
par(cex = 1.8)

curve(dnorm(x), from = -4, to= 4,
     main = "Przedział ufności wokół średniej z próby",
     ylab = 'p',
     xlab = expression(bar(X)),
     xaxt = 'n',
     type = 'n')
abline(v=qnorm(0.025), lty = 3)
abline(v=qnorm(0.975), lty = 3)

cid <- expression(bar(X) - qnorm(0.025) %*% sigma/sqrt(N))
ciq <- expression(bar(X) - qnorm(0.975) %*% sigma/sqrt(N))

lab1 <- expression(mu[d])
lab2 <- expression(mu[g])

text(x = qnorm(0.025) + 0.8, y = -0.01, labels = cid, cex = 0.5)
text(x = qnorm(0.975) - 0.8, y = -0.01, labels = ciq, cex = 0.5)

axis(side =1 , at=qnorm(c(0.025, 0.975)), labels= c(lab1, lab2))

curve(dnorm(x,qnorm(0.025)), add = TRUE)
curve(dnorm(x,qnorm(0.975)), add = TRUE)

curve(dnorm(x,qnorm(0.25)), add = TRUE, col = 'gray')
curve(dnorm(x,qnorm(0.40)), add = TRUE, col = 'gray')
curve(dnorm(x,qnorm(0.75)), add = TRUE, col = 'gray')
curve(dnorm(x,qnorm(0.90)), add = TRUE, col = 'gray')
```



W pewnym sensie konstruowanie przedziału ufności to odwrotność testowania hipotezy zerowej: określamy przedział takich możliwych wartości średniej w populacji, dla których nie odrzucilibyśmy hipotezy zerowej. 95-procentowy przedział ufności to przedział, który w 95% przypadków obejmie prawdziwą średnią w populacji. Poniższy kod losuje `k` prób z zadanej populacji i dla każdej z nich rysuje odcinek odpowiadający 95-procentowemu przedziałowi ufności wokół średniej.

```
mu <- 165; sigma <- 5; N <- 20 # parametry rozkładu, z którego będziemy losować próbę
k <- 50 # liczba losowań

par(cex = 1.8)
plot(seq(mu-15, mu+15, length.out=k), 1:k,
     type="n", xlab="", ylab="Próba nr",
     main=expression("Przedział ufności wokół średniej z próby"))

abline(v=mu)
mtext(expression(mu), 3)

for(i in 1:k) {
   sample <- rnorm(N, mu, sigma)
   segments(x0 = mean(sample) - qnorm(.975) * sigma/sqrt(N),
            x1 = mean(sample) + qnorm(.975) * sigma/sqrt(N), y0=i, lwd=5)
}
```



Oto uproszczona wersja powyższej symulacji: zamiast wykresu zwraca pojedyńczą liczbę.

Każda iteracja wyrażenia zawartego w funkcji `replicate()` zwraca wartość logiczną, `TRUE` albo `FALSE`, w zależności od tego, czy przedział ufności objął średnią w populacji, czy nie. W efekcie uzyskujemy wektor 100000 takich wartości logicznych. Funkcja `sum()` zamienia je na zera i jedynki i zwraca tym samym liczbę "trafień". Dzięki temu dowiadujemy się, ile razy przedział ufności wokół średniej w próbie objął średnią w populacji, a dzieląc tę liczbę przez 100000 dostajemy proporcję "trafień".

Pytanie

Wokół jakiej wartości oscylują liczby zwracane przez ten kod?

Odpowiedź

Wokół wartości 0.95

```
i <- 100000
sum(replicate(i, {
   sample <- rnorm(N, mu, sigma)
   mu >= mean(sample) - qnorm(.975) * sigma/sqrt(N) &
   mu <= mean(sample) + qnorm(.975) * sigma/sqrt(N)
})) / i
```

```
## [1] 0.95021
```

Pytanie

Jak wpływa na poniższe symulacje manipulacja wartościami N i σ?

- Im większe N, tym węższy przedział ufności.
- Im większa σ, tym szerszy przedział ufności. (szersze parówki)
- Im większe N, tym częściej przedział ufności obejmuje średnią w populacji.
- Im większa σ, tym częściej przedział ufności obejmuje średnią w populacji.

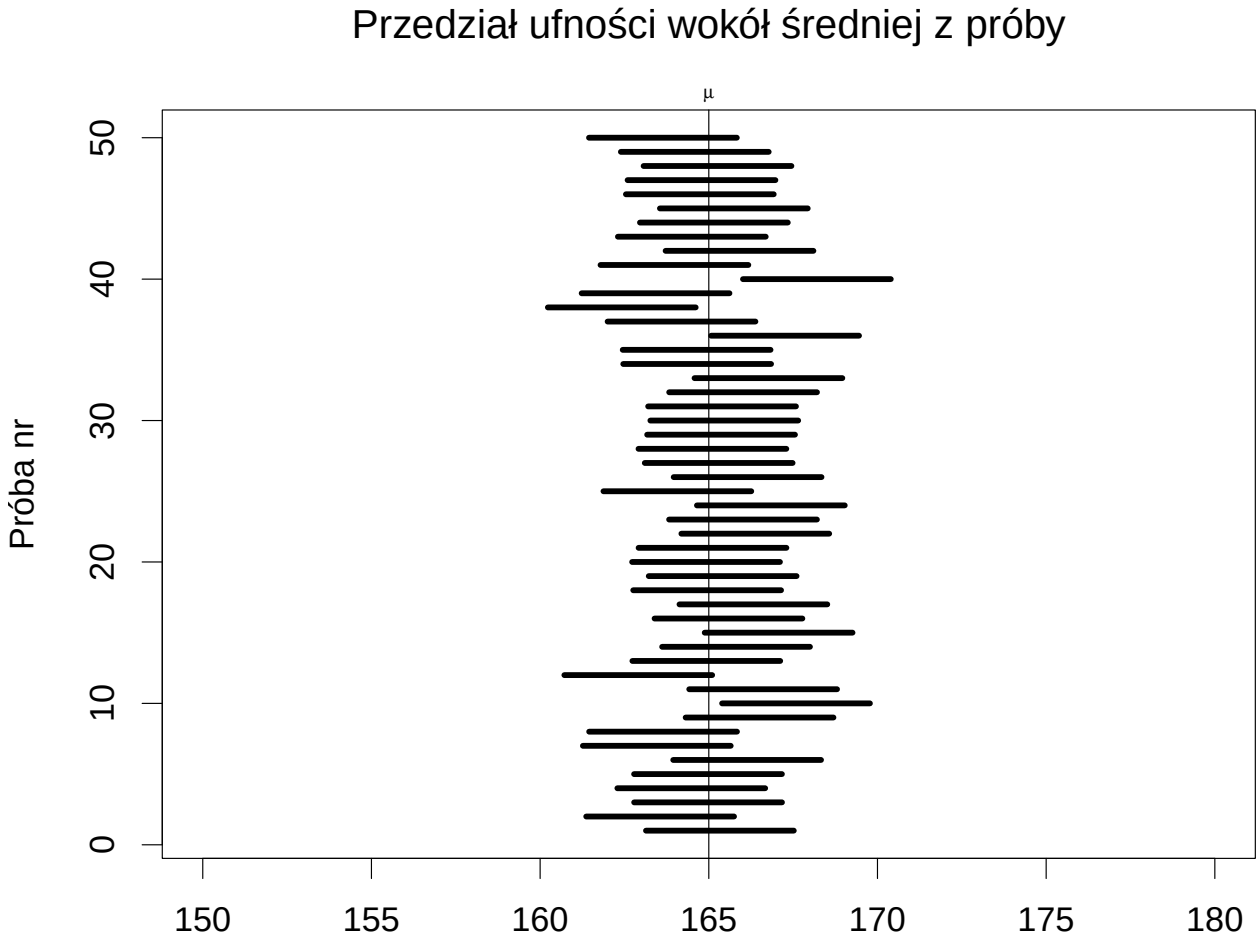
Odpowiedź

Prawidłowe są dwie pierwsze odpowiedzi:

- Im większe N, tym węższy przedział ufności.
- Im większa σ, tym szerszy przedział ufności.

Zarówno odchylenie standardowe w populacji, jak i wielkość próby, wpływają na błąd standardowy, a zatem precyzję naszej estymacji, czyli na szerokość przedziału ufności. Nie mają natomiast związku (i o to chodzi!) z poziomem ufności, który sami ustalamy arbitralnie (np. na 95%) w oparciu o rozkład normalny.

```
mu <- 165; sigma <- 5; N <- 20
k <- 50

par(cex = 1.8)
plot(seq(mu-15, mu+15, length.out=k), 1:k,
     type="n", xlab="", ylab="Próba nr",
     main=expression("Przedział ufności wokół średniej z próby"))
abline(v=mu)
mtext(expression(mu), 3)
for(i in 1:k) {
   sample <- rnorm(N, mu, sigma)
   segments(x0 = mean(sample) - qnorm(.975) * sigma/sqrt(N),
            x1 = mean(sample) + qnorm(.975) * sigma/sqrt(N), y0=i, lwd=5)
}
```




```
i <- 100000
sum(replicate(i, {
   sample <- rnorm(N, mu, sigma)
   mu >= mean(sample) - qnorm(.975) * sigma/sqrt(N) &
   mu <= mean(sample) + qnorm(.975) * sigma/sqrt(N)
})) / i
```

```
## [1] 0.94894
```

Do tej pory milcząco zakładaliśmy, że konstruując przedział ufności wokół średniej (albo testując dotyczącą jej hipotezę zerową) znamy odchylenie standardowe zmiennej w populacji. Takie sytuacje rzeczywiście się zdarzają, jednak o wiele częściej jest tak, że tego parametru wcale nie znamy. Skoro nie znamy średniej w populacji, dlaczego mielibyśmy znać odchylenie standardowe?

Pytanie

Co w takiej sytuacji? Czy nie znając odchylenia w populacji możemy w ogóle szacować jej średnią? A co się stanie, jeśli podmienimy nieznane odchylenie standardowe w populacji na odchylenie standardowe policzone w próbie? Proszę zmodyfikować odpowiednio kod symulacji i sprawdzić, co się stanie.

- Długość przedziału ufności zmienia się z próby na próbę.
- Długość przedziału ufności pozostaje taka sama z próby na próbę.
- Przedział ufności obejmuje średnią w populacji trochę częściej niż wskazywałoby jego "nominalne" 95%.
- Przedział ufności obejmuje średnią w populacji nieznacznie częściej niż w połowie przypadków.

Odpowiedź

Prawidłową odpowiedzią jest odpowiedź pierwsza.

Okazuje się, że odchylenie standardowe w próbie jest całkiem dobrym estymatorem odchylenia standardowego w populacji: przedziały ufności nadal "działają". Ponieważ jednak to odchylenie za każdym razem liczymy z próby i zmienia się ono z próby na próbę, zmienna jest również długość przedziału: jego "ramię", zamiast

`qnorm(.975) * sigma/sqrt(N)`,

wynosi

`qnorm(.975) * sd(sample)/sqrt(N)`.

Mogłoby się wydawać, że długość "ramienia" raz będzie przeszacowana, a raz niedoszacowana. Tymczasem okazuje się, że używając odchylenia standardowego z próby systematycznie zawężamy przedział, którego faktyczny poziom ufności staje się niższy niż "nominalne" 95%. Proszę wykonać jeszcze raz tę symulację, kilka razy dla $N = 10$ i kilka razy dla $N = 100$:

```
mu <- 165; sigma <- 5; N <- 10
i <- 100000
sum(replicate(i, {
   sample <- rnorm(N, mu, sigma)
   mu >= mean(sample) - qnorm(.975) * sd(sample)/sqrt(N) &
   mu <= mean(sample) + qnorm(.975) * sd(sample)/sqrt(N)
})) / i
```

```
## [1] 0.91818
```

95-procentowy przedział ufności obejmuje średnią w populacji rzadziej niż w 95000 na 100000 przypadków. Im mniejsza próba, tym jest gorzej. Dla $N = 10$ faktyczny poziom ufności spadł poniżej 92%. Dla $N = 100$ spadek jest natomiast nieznaczny.

Pytanie

Wyjaśnienie przynosi symulacja rozkładu wariancji z próby. Proszę z jej pomocą zidentyfikować prawidłową odpowiedź poniżej.

- Ponieważ rozkład wariancji z próby, zwłaszcza dla małych prób, jest prawostronnie skośny, jest większe prawdopodobieństwo, że pojedyncza wartość s niedoszacowuje σ, niż że je przeszacowuje.
- Ponieważ rozkład wariancji z próby, zwłaszcza dla małych prób, jest lewostronnie skośny, jest większe prawdopodobieństwo, że pojedyncza wartość s niedoszacowuje σ, niż że je przeszacowuje.
- Ponieważ rozkład wariancji z próby, zwłaszcza dla małych prób, jest prawostronnie skośny, jest większe prawdopodobieństwo, że pojedyncza wartość s przeszacowuje σ, niż że je niedoszacowuje.
- Ponieważ rozkład wariancji z próby, zwłaszcza dla małych prób, jest lewostronnie skośny, jest większe prawdopodobieństwo, że pojedyncza wartość s przeszacowuje σ, niż że je niedoszacowuje.

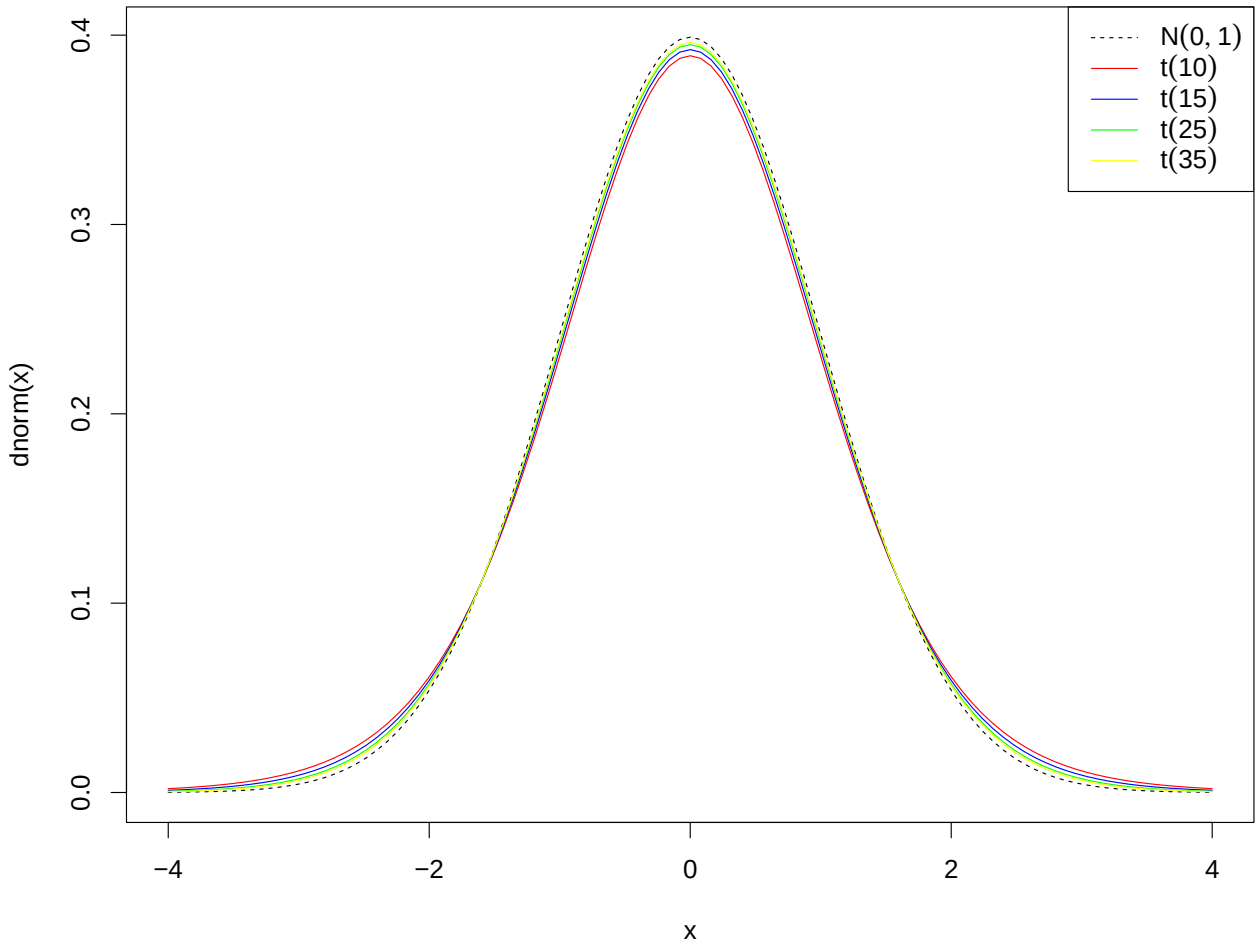
Odpowiedź

Rozkład wariancji z próby jest lekko prawoskośny, tzn. ma dłuższy ogon z prawej strony, czyli średnią wyższą od mediany, czyli większa część tego rozkładu znajduje się poniżej średniej. Zatem losując pojedyńczą próbę, mamy większe prawdopodobieństwo, że jej wariancja będzie niższa od wariancji w całej populacji. Im większa liczebność próby, tym bardziej ta skośność zanika, i przy N jeszcze poniżej 100 jest już w zasadzie zaniedbywalna.

Niemniej, kiedy nie znamy σ i opieramy się na s, od rozkładu normalnego lepszym przybliżeniem jest rozkład t. Rozkład ten wygląda podobnie do standardowego rozkładu normalnego, ale ma grubsze ogony, a zatem dalej zaczyna się obszar krytyczny i dłuższe są przedziały ufności oparte na tym rozkładzie. Jest jeden parametr, który dokładnie określa kształt rozkładu t: liczba stopni swobody. Im mniejsza próba (mniej stopni swobody), tym trudniej odrzucić hipotezę zerową i tym mniej precyzyjny (dłuższy) przedział ufności wokół średniej.

```
par(cex = 1.4)

curve(dnorm(x), from = -4, to = 4, col = 'black', lty = 2)
```

```
curve(dt(x,10), from = -4, to = 4, col = 'red', add = TRUE)
curve(dt(x,15), from = -4, to = 4, col = 'blue', add = TRUE)
curve(dt(x,25), from = -4, to = 4, col = 'green', add = TRUE)
curve(dt(x,35), from = -4, to = 4, col = 'yellow', add = TRUE)
legend('topright',
       legend = c(expression(N(0,1)),
                  expression(t(10)),
                  expression(t(15)),
                  expression(t(25)),
                  expression(t(35))),
       lty = c(2,1,1,1,1), col = c('black', 'red', 'blue', 'green', 'yellow')
      )
```



Jak pamiętamy wzór na statystykę testową z wyglądał tak:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Zmodyfikowany tak, że uwzględnia fakt, że estymujemy odchylenie standardowe wygląda tak:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$

Tak zmodyfikowana symulacja powinna z powrotem zwracać wartości oscylujące blisko .95:

```
mu <- 165; sigma <- 5; N <- 20
i <- 100000
sum(replicate(i, {
   sample <- rnorm(N, mu, sigma)
   mu >= mean(sample) - qt(.975, N-1) * sd(sample)/sqrt(N) &
   mu <= mean(sample) + qt(.975, N-1) * sd(sample)/sqrt(N)
})) / i
```

```
## [1] 0.9507
```

Zadanie

Na koniec zadanie:

Pewni amerykańscy badacze (Compas et al., 1994) przyglądali się rozwojowi dzieci dorastających w stresujących warunkach. Badacze ci z zaskoczeniem odkryli, że badane przez nich dzieci zgłaszały mniej objawów lęku czy depresji, niż się tego spodziewano. Równocześnie jednak wyniki tych dzieci na Skali Kłamstwa (narzędzie mierzące tendencję do udzielania społecznie oczekiwanych odpowiedzi) były wyższe niż przeciętna. Średni wynik na Skali Kłamstwa w amerykańskiej populacji wynosi 3.87, a w próbie 36 dzieci objętych tym badaniem średnia wyniosła 4.39, a odchylenie standardowe — 2.61.

Czy rzeczywiście badane dzieci miały podwyższoną tendencję do udzielania społecznie oczekiwanych odpowiedzi? Użyj funkcji `qt` oraz `pt`. Wybierz właściwe odpowiedzi:

- Wartość statystyki t wynosi 1.1954 i jest niższa od wartości krytycznej dla dwustronnego $\alpha = 0.05$ (2.03011). Nie mamy zatem podstaw, żeby uznać, że badane dzieci miały istotnie podwyższoną tendencję do udzielania społecznie oczekiwanych odpowiedzi.
- Dolna granica 95-procentowego przedziału ufności wokół średniej w próbie wynosi 3.5069, a zatem średnia w populacji (3.87) mieści się w tym przedziale. Nie mamy zatem podstaw, żeby uznać, że badane dzieci miały istotnie podwyższoną tendencję do udzielania społecznie oczekiwanych odpowiedzi.
- Zgodnie z hipotezą zerową ta próba 36 dzieci została wylosowana z populacji o średniej 3.87.
- Wartość statystyki t wynosi 1.1954 i jest niższa od wartości krytycznej dla dwustronnego $\alpha = 0.05$ (1.96). Nie mamy zatem podstaw, żeby uznać, że badane dzieci miały istotnie podwyższoną tendencję do udzielania społecznie oczekiwanych odpowiedzi.

Odpowiedź

Tylko ostatnia odpowiedź jest nieprawidłowa - musimy posłużyc się rozkładem t, nie zaś rozkładem normalnym.