Testowanie statystyczne hipotez

Statystyka I z R

Testowanie hipotez

Testowanie statystyczne hipotez jest jedną z najważniejszych technik statystyki inferencyjnej. Używa się jej niemal zawsze wtedy, gdy pracujemy z danymi pochodzącymi z badań eksperymentalnych lub obserwacyjnych. Istnieje bardzo wiele testów statystycznych, ale w tej części kursu skoncentrujemy się na samej logice testowania hipotez.

Procedurę statystycznego testowania hipotez możemy opisać w kilku krokach:

1. Wymyślamy naszą hipotezę alternatywną (ta nas interesuje!)
2. Przyjmujemy hipotezę zerową (i ewentualnie dodatkowe założenia).
3. Decydujemy się na określony poziom istotności α.
4. Konstruujemy rozkład interesującej nas statystyki z próby (przy założeniu H_0 i ew. innych założeniach). (Naszą statystyką z próby będzie w tym przypadku liczba sukcesów (w przypadku testu dwumianowego) i średnia z próby (w przypadku testu z))
5. Określamy region krytyczny (biorąc pod uwagę α i ew. kierunkowość hipotezy).
6. Losujemy próbę, zbieramy dane, liczymy statystykę.
7. Odrzucamy, bądź nie H_0

Hipoteza zerowa i hipoteza alternatywna

Hipoteza zerowa (H_0) - Jest to hipoteza poddana procedurze weryfikacyjnej, w której zakładamy, że różnica między analizowanymi parametrami lub rozkładami wynosi zero (nie zawsze nasza hipoteza zerowa głosi, że coś wynosi 0, ale jest tak w większosci wypadków).

Przykładowo, wnioskując o parametrach hipotezę zerową zapiszemy jako:

$$H_0 : \theta_1 = \theta_2$$

gdzie θ_1 i θ_2 to parametry rozkładu populacji - na przykład średnie albo odchylenia standardowe.

Hipoteza alternatywna (H_1 lub H_A) - hipoteza przeciwstawna do weryfikowanej. Możemy ją zapisać na trzy sposoby w zależności od sformułowania badanego problemu:

Hipoteza dwustronna (wartości dwóch parametrów nie są równe)

$$H_1 : \theta_1 \neq \theta_2$$

Hipoteza prawostronna (wartość pierwszego parametru jest większa niż drugiego)

$$H_1 : \theta_1 > \theta_2$$

Hipoteza lewostronna (wartość pierwszego parametru jest mniejsza niż drugiego)

$$H_1 : \theta_1 < \theta_2$$

W zależności od przyjętej hipotezy alternatywnej będziemy przeprowadzać sam test statystyczny odrobinę inaczej.

Błąd typu I i błąd typu II

Błąd pierwszego rodzaju (błąd pierwszego typu, alfa-błąd, ang. *false positive*) – błąd polegający na odrzuceniu hipotezy zerowej, która w rzeczywistości nie jest fałszywa. Oszacowanie prawdopodobieństwa popełnienia błędu pierwszego rodzaju oznaczamy symbolem α (mała grecka litera alfa) i nazywamy poziomem istotności testu. My będziemy na zajęciach z reguły przyjmować poziom $\alpha = 0.05$, ale coraz częściej w literaturze podnosi się, że powinniśmy być w naszych analizach bardziej restrykcyjni i przyjąć $\alpha = 0.01$ lub nawet $\alpha = 0.001$.

Błąd drugiego rodzaju (błąd drugiego typu, błąd przyjęcia, beta-błąd, ang. *false negative*) – błąd polegający na nieodrzuceniu hipotezy zerowej, która jest w rzeczywistości fałszywa. Prawdopodobieństwo popełnienia błędu drugiego rodzaju często oznaczamy symbolem β (mała grecka litera beta). Często w kontekście błędu drugiego rodzaju będziemy mówić o mocy testu statystycznego, która wynosi $1 - \beta$.

P-wartość, prawdopodobieństwo testowe (ang. p-value, *probability value*), graniczny poziom istotności – prawdopodobieństwo uzyskania wartości pewnej statystyki (np. różnicy średnich) takich, jak faktycznie zaobserwowane, lub bardziej oddalonych od zera, przy założeniu, że hipoteza zerowa jest spełniona. Stosowane jako miara prawdopodobieństwa popełnienia błędu pierwszego rodzaju, czyli liczbowe wyrażenie istotności statystycznej.

(źródło: Wikipedia)

Testowanie hipotez: przykład I

Wyobraźmy sobie, że interesuje nas prawdopodobieństwo dostania astmy na przestrzeni roku wśród dzieci w wieku od 0 do 4 lat, których matka pali w domu papierosy.

Skądinąd wiemy, że rocznie około 1.4% dzieci w wieku do 4 lat dostaje astmy.

Zakładamy, że palenie w domu raczej nie zmniejsza ryzyka astmy, tylko zwiększa, więc naszą hipotezą alternatywną jest hipoteza prawostronna:

$$H_A : p > 0.014$$

Nasza hipoteza zerowa, którą będziemy testować będzie jej negacją:

$$H_0 : p = 0.014$$

Badanie wykazało, że w próbie 500 dzieci z populacji dzieci, których matki palą w domu, 10 dostało astmy. Czy uprawnia nas to do odrzucenia hipotezy zerowej?

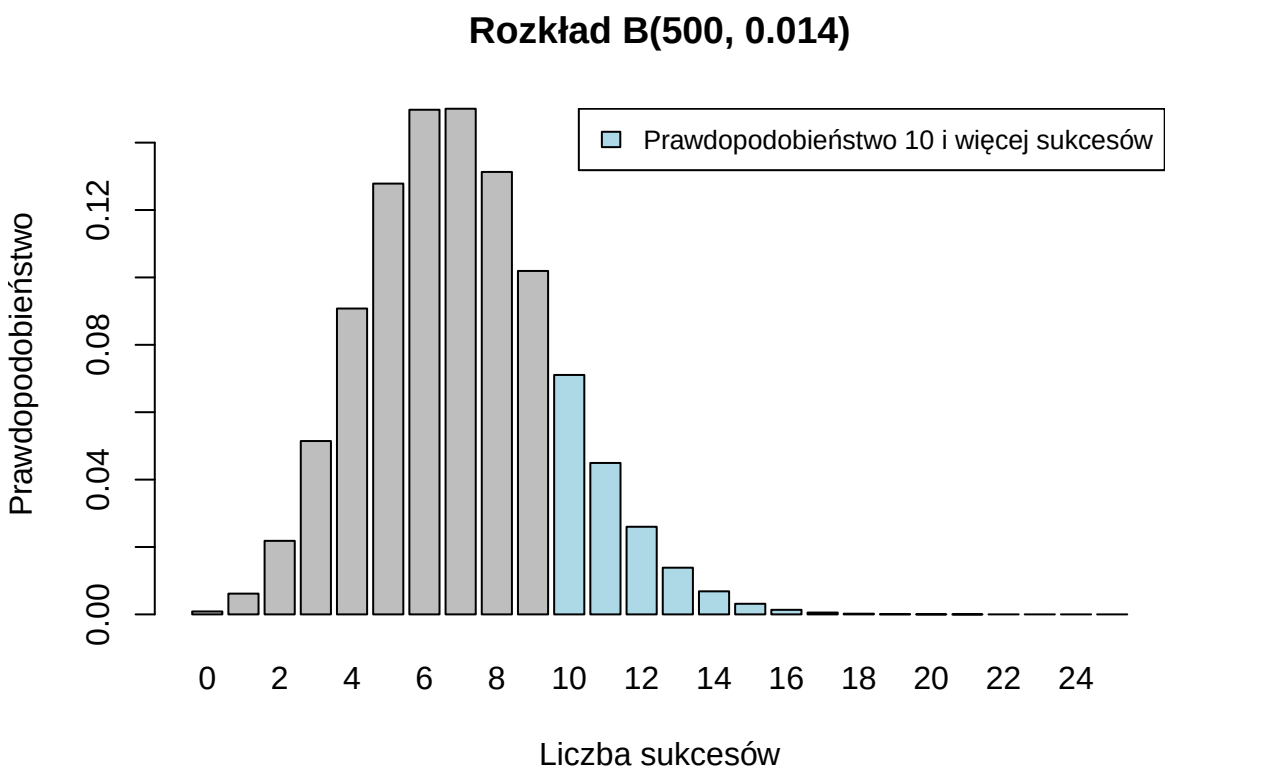
Interesuje nas następujące pytanie: jakie jest prawdopodobieństwo tego, że uzyskaliśmy taki wynik, jaki uzyskaliśmy, oczywiście zakładając, że hipoteza zerowa jest prawdziwa.

Problem, który rozważamy jest problemem dwumianowym, więc możemy dokładnie to prawdopodobieństwo poznać.

Załóżmy (zgodnie z hipotezą zerową), że prawdopodobieństwo sukcesu wynosi 0.14. Interesuje nas więc jakie jest prawdopodobieństwo 10 i więcej sukcesów przy liczbie prób wynoszącej 500. Możemy zobrazować to na prostym wykresie:

```
# za pomocą `dbinom` tworzymy wektor z prawdopodobieństwami 0,1,2,...,25 sukcesów
barplot(dbinom(0:25, 500, 0.014),
        names.arg = 0:25, # nazwy słupków
        col=c(rep('grey', 10), rep('lightblue', 15)),
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',
        main = 'Rozkład B(500, 0.014)') # sposób pokolorowania słupków
```

```r
# uwaga, stworzyłem wykres obejmujący od 0 do 25 sukcesów
# tylko ze względu na jasność prezentacji;
# w rzeczywistości oczywiście jest możliwe nawet 500 sukcesów

legend('topright',
       fill = 'lightblue',
       legend = 'Prawdopodobieństwo 10 i więcej sukcesów',
       cex = 0.8 # zmniejszymy czcionkę dla czytelności
       )
```




```r
# kiedy używamy `lower.tail = FALSE` musimy odjąć 1,
# ponieważ dystrybuanta bierze pod uwagę nieostrą nierówność
print('Prawdopodobieństwo 10 i więcej sukcesów:')
## [1] "Prawdopodobieństwo 10 i więcej sukcesów:"
print(pbinom(10-1,500, 0.014, lower.tail = FALSE))
## [1] 0.1680696
```

Interesuje nas tylko prawy ogon tego rozkładu ze względu na kierunkowość naszej hipotezy alternatywnej. Możemy korzystając z dystrybuanty z funkcji `pbinom` stwierdzić, że prawdopodobieństwo 10 lub więcej sukcesów wynosi ≈ 0.168. Nie daje nam to powodu do odrzucenia hipotezy zerowej, jeżeli wybraliśmy poziom istotności statystycznej $\alpha = 0.05$

Testowanie hipotez: przykład II

Interesują nas zdolności poznawcze dzieci o wadze urodzeniowej mniejszej niż 1500 gramów, które cierpiały na zaburzenia rozwoju płodu. W populacji dzieci rozwijających się normalnie, 5% ma iloraz inteligencji niższy niż 70, gdy osiagają wiek 8 lat. Czy jest to prawdą także dla dzieci, które cierpią na zaburzenia rozwoju płodu? Aby to sprawdzić, wybrano losową próbę 66 dzieci, wśród których, jak się okazało, 7 miało iloraz inteligencji poniżej 70 punktów.

Nasza hipoteza alternatywna głosi, że prawdopodobieństwo uzyskania ilorazu niższego niż 70 punktów w wieku 8 lat jest wśród dzieci z zaburzeniami rozwoju płodu wyższe niż w ogólnej populacji (5%).

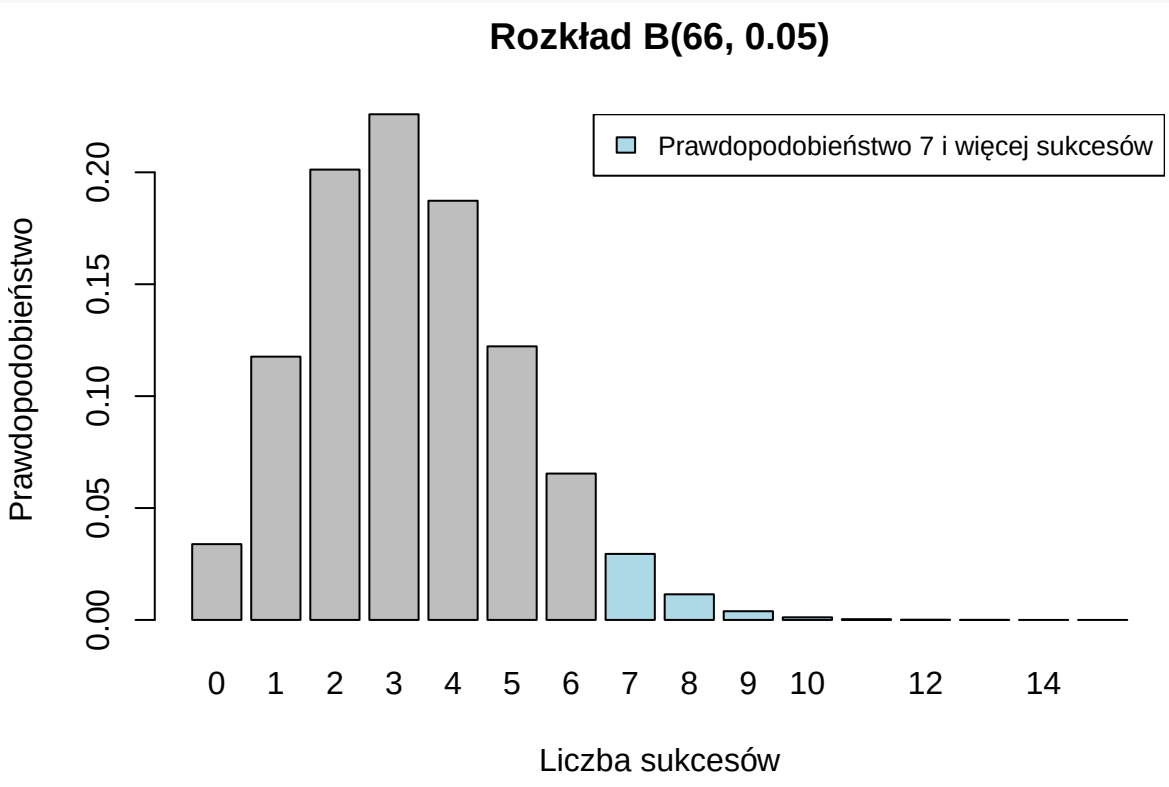
$$H_A : p > 0.05$$

Hipoteza zerowa głosi więc, że prawdopodobieństwo jest takie samo, jak w ogólnej populacji.

$$H_0 : p = 0.05$$

Podobnie jak w poprzednim przykładzie możemy zobrazować prawdopodobieństwo, które nas interesuje, za pomocą wykresu, a następnie obliczyć dokładną wartość prawdopodobieństwa testowego, używając funkcji `pbinom`.

```
barplot(dbinom(0:15, 66, 0.05),
        names.arg = 0:15,
        col=c(rep('grey', 7), rep('lightblue', 8)),
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',
        main = 'Rozkład B(66, 0.05)')


legend('topright',
       fill = 'lightblue',
       legend = 'Prawdopodobieństwo 7 i więcej sukcesów',
       cex = 0.8
       )
```




```
print('Prawdopodobieństwo 7 i więcej sukcesów:')
## [1] "Prawdopodobieństwo 7 i więcej sukcesów:"
```

```r
print(pbinom(7-1, 66, 0.05, lower.tail = FALSE))
## [1] 0.04641626
```

Tym razem prawdopodobieństwo uzyskania 7 lub więcej sukcesów przy założeniu prawdziwości hipotezy zerowej wynosi 0.046. Wartość ta uprawnia nas do odrzucenia hipotezy zerowej i (jak sądzą niektórzy) do przyjęcie hipotezy alternatywnej, która głosi, że prawdopodobieństwo uzyskania ilorazu inteligencji niższego niż 70 wśród ośmiolatków o wadze urodzeniowej poniżej 1500 gramów cierpiących na zaburzenia rozwoju płodu jest wyższe od prawdopodobieństwa takiego wyniku wśród zdrowych ośmiolatków.

Testowanie hipotez: przykład III - *Bond tasting Martini*

James Bond słynny jest z tego, że zawsze zamawia Martini wstrząśnięte, nie zmieszane. Czy jednak rzeczywiście potrafi odróżnić te dwie metody przygotowania swojego ulubionego drinku? Q postanowił go sprawdzić i zaprojektował prosty eksperyment. Podczas 16 losowych prób Bond miał orzec, czy Martini, które właśnie pił, było wstrząśnięte, czy zmieszane. Gdyby wybierał całkowicie losowo, miałby 50% szans na poprawne zgadnięcie.

Eksperyment pokazał, że Bond zgadł w 13 próbach na 16.

1. Czy uprawnia nas to do twierdzenia, że odpowiedział lepiej, niż gdyby odpowiadał losowo?

Nasza hipoteza alternatywna głosi, że prawdopodobieństwo sukcesu (czyli prawidłowego odgadnięcia) jest wyższe, niż gdyby odpowiadał całkowicie losowo (50%).

$$H_A : p > 0.50$$

Hipoteza zerowa głosi więc, że Bond nie jest istotnie lepszy, niż gdyby zgadywał.

$$H_0 : p = 0.50$$

2. Możemy jednak sformułować nasze pytanie badawcze nieco inaczej. Być może nie chcemy od razu rozstrzygać, że Bond jest lepszy, niż gdyby odpowiadał losowo. Możemy zrezygnować z kierunkowości hipotezy alternatywnej, to znaczy zapytać, czy uzyskany przez Bonda wynik uprawnia nas to do twierdzenia, że odpowiadał *lepiej lub gorzej*, niż gdyby odpowiadał losowo? Przy takim pytanie hipoteza zerowa jest taka sama, ale hipoteza alternatywna jest nieco inna:

$$H_A : p \neq 0.50$$

$$H_0 : p = 0.50$$

W zależności od tego jak sformułujemy hipotezę alternatywną (od jej kierunkowości) zależy to, które obszary nas będą interesowały.

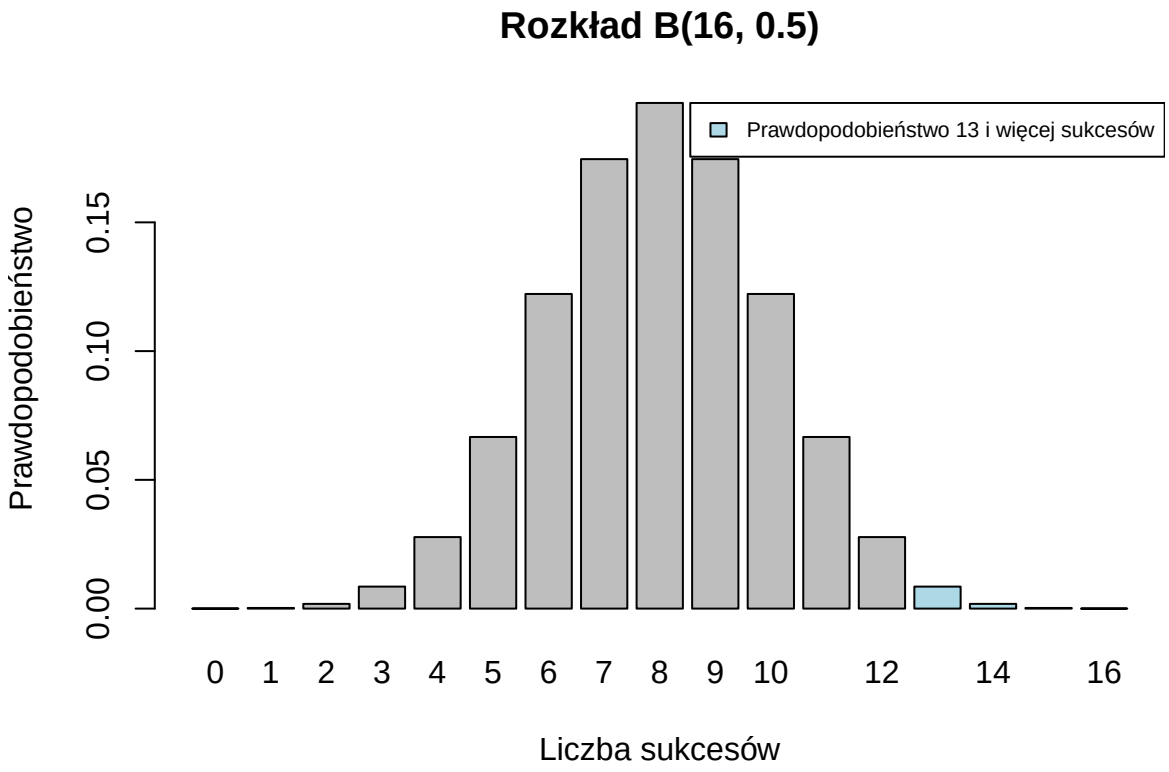
Spróbujmy odpowiedzieć na pierwsze z tych pytań, używajac R.

1. Czy uzyskany przez Bonda wynik uprawnia nas do twierdzenia, że odpowiedział lepiej, niż gdyby odpowiadał losowo?

$$H_A : p > 0.50$$

$$H_0 : p = 0.50$$

```r
barplot(dbinom(0:16, 16, 0.5),
        names.arg = 0:16,
        col=c(rep('grey', 13), rep('lightblue', 3)),
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',
        main = 'Rozkład B(16, 0.5)')

legend('topright',
       fill = 'lightblue',
       legend = 'Prawdopodobieństwo 13 i więcej sukcesów',
       cex = 0.7
       )
```




```r
print('Prawdopodobieństwo 13 i więcej sukcesów')
## [1] "Prawdopodobieństwo 13 i więcej sukcesów"
print(pbinom(12,16, 0.5, lower.tail = FALSE))
## [1] 0.01063538
```

binom.test Odpowiedni test możemy przeprowadzać nie tylko wykorzystując dystrybuantę rozkładu dwumianowego. Żeby ułatwić sobie pracę, możemy posłużyć się wbudowaną w R funkcją `binom.test`. Serdecznie rekomenduję sprawdzanie uzyskanych przez siebie wyników (szczególnie w pracach domowych) za pomocą tej funkcji.

```r
# argument `alternative` określa kierunkowość hipotezy alterantywnej.
binom.test(c(13,3), alternative = 'greater')
```

```
##
##  Exact binomial test
##
## data:  c(13, 3)
```

```
## number of successes = 13, number of trials = 16, p-value = 0.01064
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.5834277 1.0000000
## sample estimates:
## probability of success
##                 0.8125
```

Spróbujmy teraz odpowiedzieć na drugie z naszych pytań:

2. Czy uzyskany przez Bonda wynik uprawnia nas do twierdzenia, że odpowiadał lepiej lub gorzej, niż gdyby odpowiadał losowo?

$H_A : p \neq 0.50$

$H_0 : p = 0.50$

W przypadku takiego sformułowania hipotezy alternatywnej przyjmujemy, że jest on dwustronna. Nie wiemy, czy Bond odpowiada lepiej, czy gorzej, niż gdyby miał po prostu wybierać losowo. Przyjmiemy, że hipoteza zerowa została sfalsyfikowana zarówno wówczas, gdy odpowiada gorzej, niż gdyby zgadywał, jak i wtedy, gdy odpowiada lepiej. Dlatego tym razem interesują nas dwa ogony rozkładu.

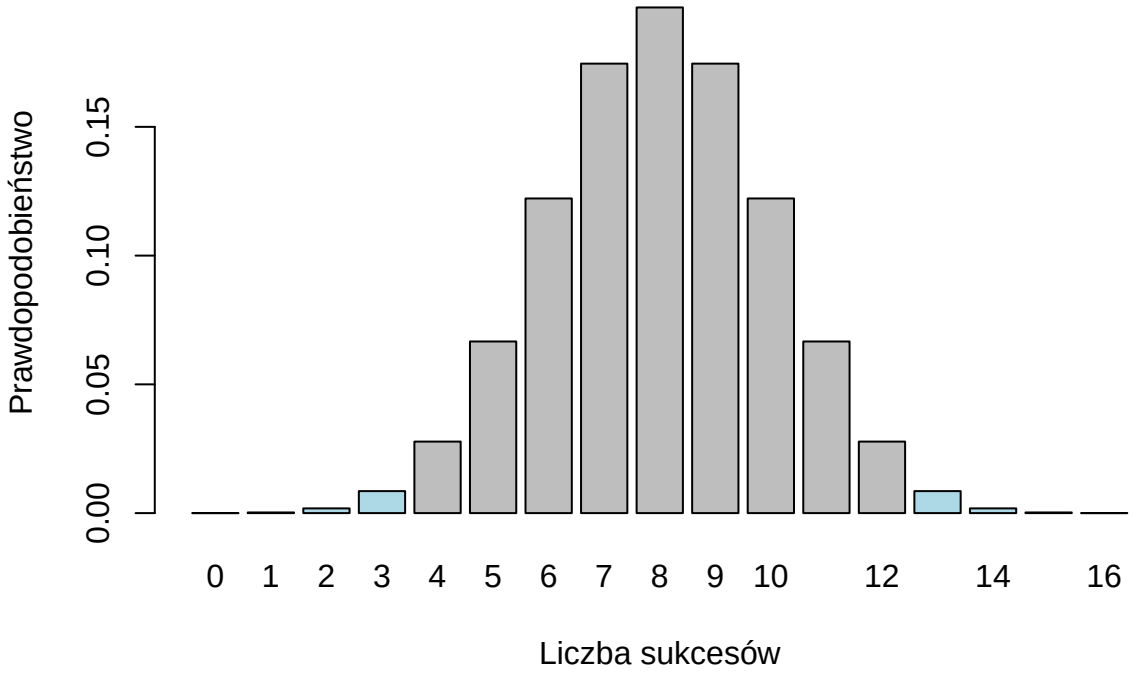
Przypomnijmy sobie, że wybrany przez nas domyślny poziom istotności statystycznej wynosi 5%. Oznacza to, że 5% "skrajnych" wyników eksperymentu (przy założeniu hipotezy zerowej!) uprawni nas do odrzucenia hipotezy zerowej. W przypadku alternatywnych hipotez kierunkowych nie jest to problemem. Co jednak, gdy hipoteza jest niekierunkowa (dwustronna)? W takim wypadku dzielimy nasze 5% na dwie cześci i sprawdzamy, czy uzyskany wynik wpada w 2.5% dolnego bądź w 2.5% górnego ogonu rozkładu.

Wyznaczając obszar odrzucenia musimy wziać pod uwagę dwustronność naszej hipotezy alternatywnej.

```
# obszar krytyczny będzie zbiorem wartości mniejszych niż 2,5 percentyl
# i większych niż 97,5 percentyl rozkładu
# jeśli uzyskana przez nas wartość leży w obszarze odrzucenia, odrzucamy hipotezę
print('Wartości krytyczne')
## [1] "Wartości krytyczne"
qbinom(c(0.025, 0.975), 16,0.5)
## [1]  4 12
```

Ponownie możemy pokazać na wykresie, jakie wyniki będą stanowiły uzysadanienie dla odrzucenia hipotezy zerowej.

```
barplot(dbinom(0:16, 16, 0.5),
        names.arg = 0:16,
        col=c(rep('lightblue',4),rep('grey', 9), rep('lightblue', 4)),
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',
        main = 'Rozkład B(16,0.5)')
```



Jak widzimy uzyskany przez Bonda wynik (13 sukcesów) znajduje się w regionie odrzucenia. Oznacza to, że również w tym przypadku odrzucamy hipotezę zerową. Aby obliczyć wartość p dla takiego testu, najczęściej obliczamy wartość p dla obu wariantów kierunkowości hipotezy alternatywnej. Następnie bierzemy mniejszą z nich i mnożymy ją przez 2.

```
# bierzemy mniejszą wartość i mnożymy przez 2
min(
  pbinom(13-1, 16, 0.5, lower.tail = FALSE),
  pbinom(13, 16,0.5)
  )*2
```

```
## [1] 0.02127075
```

```
# Możemy też skorzystać z funkcji `binom.test`
binom.test(c(13,3), p=0.5, 'two.sided')
```

```
##
##  Exact binomial test
##
## data:  c(13, 3)
## number of successes = 13, number of trials = 16, p-value = 0.02127
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5435435 0.9595263
## sample estimates:
## probability of success
##                 0.8125
```

Czas na małą dygresję. Na pewno domyślają się Państwo, w jaki sposób możemy oszacować prawdopdoobieństwo odgadnięcia przez Bonda sposobu przygotowania Martini. Część z Państwa może jednak zastanawiać się, jak skonstruować wokół takiego prawdopodobieństwa przedział ufności. Często robi się to, używając rozkładu normalnego, który jest dobrą aproksymacją rozkładu dwumianowego. Istnieją jednak lepsze sposoby.

Poniżej (dla zainteresowanych) kilka możliwości.

```
# szacowane z próby prawdopodobieństwo
print('Prawdopodobieństwo oszacowane z próby: \n')
```

Szacowane z próby prawdopodobieństwo sukcesu

```
## [1] "Prawdopodobieństwo oszacowane z próby: \n"
```

```
prob = 13/(13+3)
```

Szacowany z próby przedział ufności (z rozkładu normalnego) Ten sposób jest bardzo łatwy, ale niedokładny.

```
# z rozkładu normalnego
print('Przedział ufności oszacowany z rozkładu normalnego')
## [1] "Przedział ufności oszacowany z rozkładu normalnego"
(qnorm(c(0.025,0.975), prob, sqrt(prob*(1-prob)/16))) # sposób łatwy ale niedokładny
## [1] 0.6212505 1.0037495
```

Szacowany z próby przedział ufności (z rozkładu β) Ten sposób jest nieco lepszy.

```
print('Przedziały ufności oszacowane z rozkładu beta')
## [1] "Przedziały ufności oszacowane z rozkładu beta"
binom.test(
  matrix(c(13,3), nrow = 1),
  p= 0.5) # sposób trudny i straszny
##
##  Exact binomial test
##
## data:  matrix(c(13, 3), nrow = 1)
## number of successes = 13, number of trials = 16, p-value = 0.02127
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5435435 0.9595263
## sample estimates:
## probability of success
##                 0.8125
```

```
# jak można zobaczyć, sposób ten jest dużo dokładniejszy
print('Sprawdzenie, czy rzeczywiście tak jest')
## [1] "Sprawdzenie, czy rzeczywiście tak jest"
binom.test(c(13,3), p=0.5435435, alternative = 'g')
##
##  Exact binomial test
##
## data:  c(13, 3)
## number of successes = 13, number of trials = 16, p-value = 0.025
## alternative hypothesis: true probability of success is greater than 0.5435435
## 95 percent confidence interval:
##  0.5834277 1.0000000
## sample estimates:
## probability of success
##                 0.8125
binom.test(c(13,3), p=0.9595263, alternative = 'l')
```

```
##
##  Exact binomial test
##
## data:  c(13, 3)
## number of successes = 13, number of trials = 16, p-value = 0.025
## alternative hypothesis: true probability of success is less than 0.9595263
## 95 percent confidence interval:
##  0.000000 0.946854
## sample estimates:
## probability of success
##                 0.8125
```

```
print('Przedział ufności oszacowany za pomocą symulacji')
## [1] "Przedział ufności oszacowany za pomocą symulacji"
a = rbinom(1000000,16,prob)/16
quantile(a, c(0.025,0.975))
##  2.5% 97.5%
## 0.625 1.000
```

Szacowany z próby przedział ufności (za pomocą symulacji Monte Carlo)

Szacowany z próby przedział ufności (za pomocą funkcji `binom.confint` z pakietu `binom`) Możemy też posłużyć się pakietem `binom`, w którym znajduje się funkcja `binom.confint` oferująca różne inne metody wyznaczania przedziałów ufności.

```
library(binom)
binom.confint(13,16)
##           method  x  n      mean     lower     upper
## 1  agresti-coull 13 16 0.8125000 0.5619784 0.9420168
## 2     asymptotic 13 16 0.8125000 0.6212505 1.0037495
## 3          bayes 13 16 0.7941176 0.6073333 0.9602614
## 4        cloglog 13 16 0.8125000 0.5246043 0.9353523
## 5          exact 13 16 0.8125000 0.5435435 0.9595263
## 6          logit 13 16 0.8125000 0.5525441 0.9382961
## 7         probit 13 16 0.8125000 0.5700893 0.9449442
## 8        profile 13 16 0.8125000 0.5827611 0.9497199
## 9            lrt 13 16 0.8125000 0.5827632 0.9497621
## 10     prop.test 13 16 0.8125000 0.5369234 0.9502687
## 11        wilson 13 16 0.8125000 0.5699112 0.9340840
```

Testowanie hipotez - wartość p przy niesymetrycznych rozkładach

Załóżmy, że przeciętny student ma 80% szans na zdanie egzaminu. Wybraliśmy próbę 50 studentów kognitywistyki, przeegzaminowaliśmy i okazało się, że 35 z nich zdało egzamin. Czy studenci kognitywistyki są lepsi lub gorsi od reszty studentów?

Ze względu na to, że nasze pytanie badawcze nie specyfikuje, że chcieliśmy testować hipotezę kierunkową. Układ hipotez przedstawia się więc w sposób następujący:

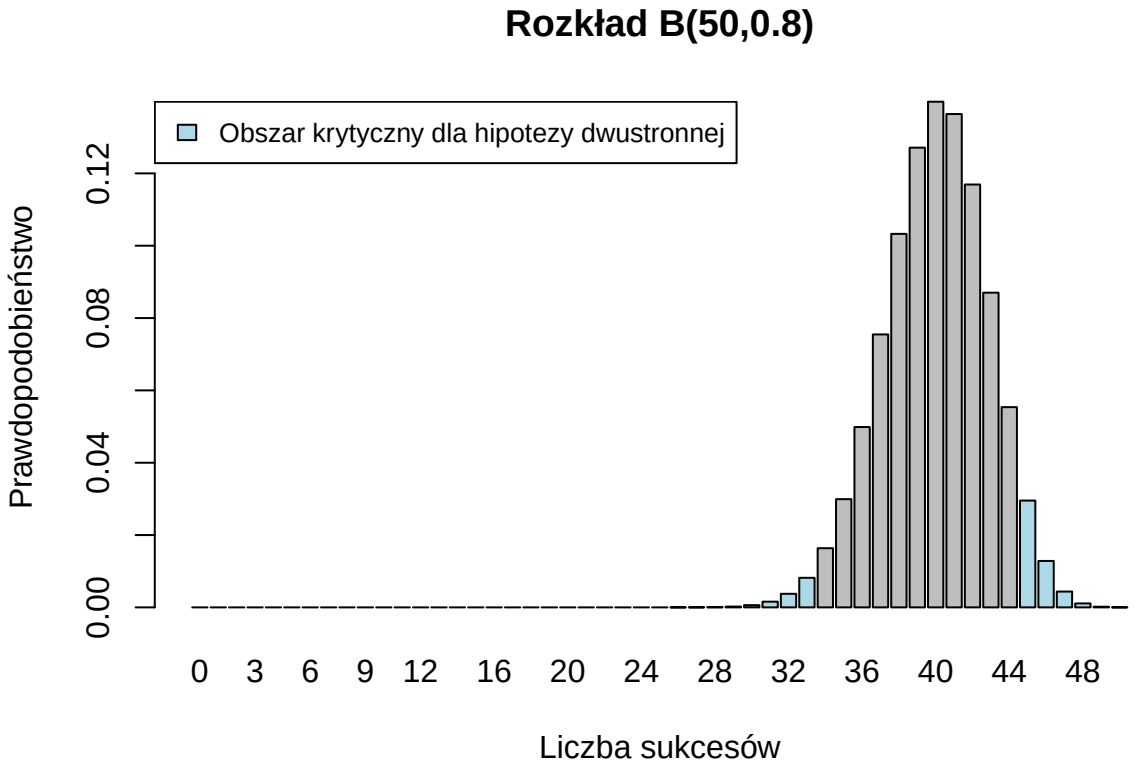
$$H_A : p \neq 0.80$$

$$H_0 : p = 0.80$$

Na początek przedstawmy obszary krytyczne na wykresie.

```
alpha <- 0.05
obszar_odrzucenia <- qbinom(c(alpha/2, 1- (alpha/2)),
                            50,
                            0.8)
print('Wartości krytyczne')
## [1] "Wartości krytyczne"
print(obszar_odrzucenia)
## [1] 34 45
```

```
barplot(dbinom(0:50, 50, 0.8),
        names.arg = 0:50,
        col=c(rep('lightblue',obszar_odrzucenia[1]),
              rep('grey',50 - obszar_odrzucenia[1] - (50 - obszar_odrzucenia[2])),
              rep('lightblue', 50-obszar_odrzucenia[2])),
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',
        main = 'Rozkład B(50,0.8)'
        )

legend('topleft',
       fill = 'lightblue',
       legend = 'Obszar krytyczny dla hipotezy dwustronnej',
       cex = 0.8)
```



Jak widzimy uzyskana przez nas liczba studentów, którzy zdali egzamin nie wpada w żaden z dwóch obszarów odrzucenia. Spróbujmy jednak obliczyć dokładną wartość prawdopodobieństwa testowego (*p value*). Najpierw obliczymy wartość dla testu lewostronnego, potem dla prawostronnego.

```r
print('Prawdopodobieństwo 35 i mniej sukcesów')
## [1] "Prawdopodobieństwo 35 i mniej sukcesów"
pbinom(35, 50, 0.8)
## [1] 0.06072208
print('Prawdopodobieństwo 35 i więcej sukcesów')
## [1] "Prawdopodobieństwo 35 i więcej sukcesów"
pbinom(35-1, 50, 0.8, lower.tail = FALSE)
## [1] 0.9691966
```

```r
# hipoteza lewostronna - odpowiada pbinom(35, 50, 0.8)
binom.test(c(35,15), p=0.8, alternative = 'l')
```

```
##
##  Exact binomial test
##
## data:  c(35, 15)
## number of successes = 35, number of trials = 50, p-value = 0.06072
## alternative hypothesis: true probability of success is less than 0.8
## 95 percent confidence interval:
##  0.0000000 0.8051151
## sample estimates:
## probability of success
##                    0.7
```

```r
# hipoteza prawostronna - odpowiada pbinom(35-1, 50, 0.8, lower.tail = FALSE)
binom.test(c(35,15), p=0.8, alternative = 'g')
```

```
##
##  Exact binomial test
##
## data:  c(35, 15)
## number of successes = 35, number of trials = 50, p-value = 0.9692
## alternative hypothesis: true probability of success is greater than 0.8
## 95 percent confidence interval:
##  0.576267 1.000000
## sample estimates:
## probability of success
##                    0.7
```

Wydawałoby się, że powinniśmy po prostu wziąć mniejszą z nich i następnie pomnożyć ją przez 2. Takie rozwiązanie nie zgadza się jednak z wartością podawaną przez funkcje `binom.test`, której raczej powinniśmy zaufać:

```r
# hipoteza dwustronna
binom.test(c(35,15), p=0.8, alternative = 't')
```

```
##
##  Exact binomial test
##
## data:  c(35, 15)
## number of successes = 35, number of trials = 50, p-value = 0.1087
## alternative hypothesis: true probability of success is not equal to 0.8
## 95 percent confidence interval:
##  0.5539177 0.8213822
## sample estimates:
## probability of success
```

```
##                     0.7
```

Dlaczego nasze wyniki są różne? Przy obliczaniu wartości p dla testu dwumianowego przy dwustronnej hipotezie interpretujemy "wartości bardziej skrajne niż uzyskana wartość" jako "wartości, których prawdopodobieństwo wystąpienia jest niższe, niż wartości uzyskanej w rzeczywistości".

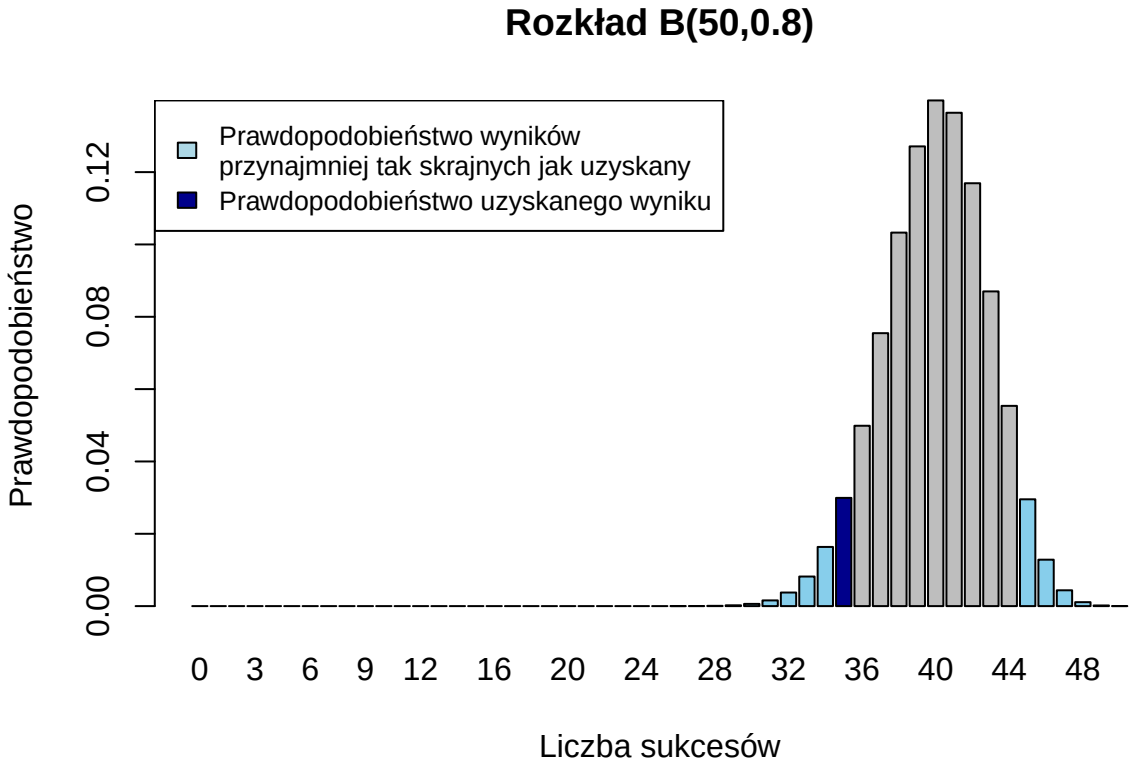
Obliczenie właściwego wyniku w R "na piechotę" wymaga kilku kroków, ale jest do zrobienia.

```
# Prawdopodobieństwo takiego wyniku, jaki otrzymalismy
d_e <- dbinom(35, 50, 0.8)

# Prawdopodobieństwa wszystkich możliwych wyników
d_w <- dbinom(0:50, 50, 0.8)

# Suma prawdopodobieństw wszystkich wyników bardziej skrajnych niż nasz (w obie strony)
sum(d_w[d_w<=d_e])
```

```
## [1] 0.1087493
```

Umiemy więc obliczyć p dla testu dwustronnego! Żeby lepiej zobrazować, o co tutaj tak naprawdę chodzi, możemy przedstawić naszą sytuację graficznie:

```
kolorki = ifelse(dbinom(0:50,50,0.8) <= d_e, 'skyblue', 'grey')
kolorki = ifelse(dbinom(0:50,50,0.8) == d_e, 'darkblue', kolorki)
b1 = barplot(dbinom(0:50, 50, 0.8),
        names.arg = 0:50,
        col = kolorki,
        main = 'Rozkład B(50,0.8)',
        xlab = 'Liczba sukcesów',
        ylab = 'Prawdopodobieństwo',)

legend('topleft', fill = c('lightblue', 'darkblue'),
       legend = c('Prawdopodobieństwo wyników \nprzynajmniej tak skrajnych jak uzyskany',
                  'Prawdopodobieństwo uzyskanego wyniku'),
       cex = 0.8)
```



Jak widzimy uzyskany wynik nie uprawnia nas do odrzucenia hipotezy zerowej. W przypadku hipotezy jednostronnej znajdujemy się "na granicy istotności statystycznej", a przypadku hipotezy dwustronnej nie odrzucilibyśmy hipotezy zerowej nawet dla $\alpha = 0.1$.

Błąd standardowy średniej i centralne twierdzenie graniczne

Błąd standardowy – odchylenie standardowe średnich z prób (gdybyśmy wiele razy pobierali próby tej samej wielkości z tej samej populacji generalnej, liczyli z nich średnie, a potem odchylenie standardowe tych średnich).

Wzór na błąd standardowy to (jeśli znamy parametry rozkładu):

$$SE = \frac{\sigma}{\sqrt{n}}$$

gdzie σ to odchylenie standardowe rozkładu populacji a n to liczebność próby

(za http://www.eko.uj.edu.pl/statystyka/II/slowniczek.pdf)

Błąd standardowy średniej wiąże się z Centralnym Twierdzeniem Granicznym, które opisuje jeden z najważniejszych faktów wykorzystywanych przy statystycznym testowaniu hipotez.

Centralne twierdzenie graniczne (jedno ze sformułowań) - Dla danej populacji o średniej μ i wariancji σ^2 rozkład średniej z próby będzie miał średnią równą μ ($\mu_{\bar{X}} = \mu$), wariancje ($\sigma^2_{\bar{X}}$) równą σ^2/n i odchylenie standardowe $\sigma_{\bar{X}}$ równe $\sigma/\sqrt{n}$.

Możemy przetestować "prawdziwość" tego twierdzenia, tworząc kilka prostych symulacji. Poniższa funkcja:

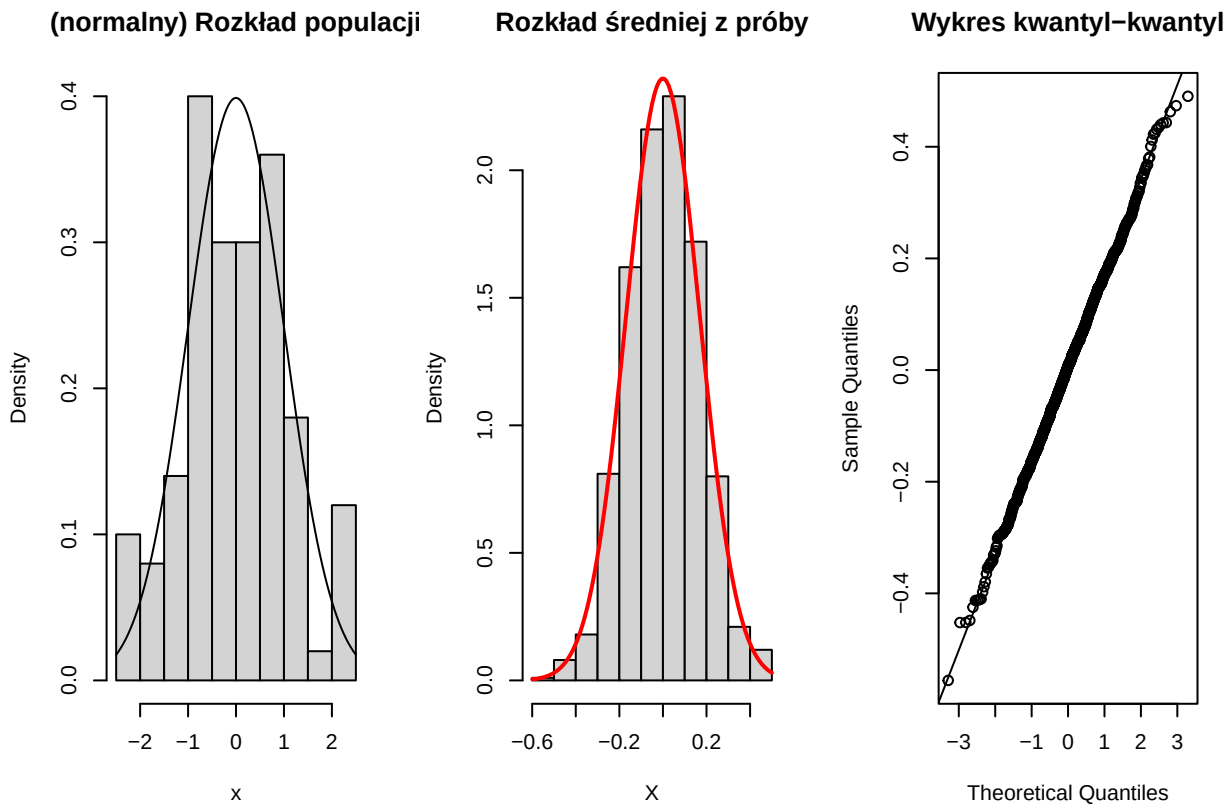
1. Z określonego rozkładu normalnego o średniej `mean` i odchyleniu standardowym `sd` losuje `i` `n`-elementowych próbek.
2. Dla każdej wylosowanej próbki oblicza średnią.
3. Dla `i` średnich tworzy trzy ilustracje:

- histogram obrazujący rozkład wartości w losowej próbie 100 obserwacji wraz z odpowiadającą mu krzywą rozkładu normalnego o średniej `mean` i odchyleniu standardowym `sd`
- histogram obrazujący rozkład średnich z `i` próbek wraz z odpowiadającą mu krzywą rozkładu normalnego o parametrach obliczonych z Centralnego Twierdzenia Granicznego, to znaczy o średniej `mean` i odchyleniu standardowym `sd/sqrt(n)`
- wykres kwantyl-kwantyl dla rozkładu średnich z `i` próbek, który pozwala ocenić, czy rozkład ten jest normalny

```
blad_standardowy <- function(n, mean=0, sd=1, i) {
  # losujemy i razy próbę o liczebności n z której liczymy średnią
  errors = replicate(i, mean(rnorm(n,mean,sd)))
  par(mfrow=c(1,3)) # chcemy narysować 3-panelowy wykres
  # Pierwszy wykres: histogram rozkładu obserwacji w próbie
  hist(
    rnorm(100, mean, sd), # losujemy 100 obserwacji
    freq = FALSE,
    main = '(normalny) Rozkład populacji',
    xlab = 'x') # rozkład empiryczny
  # Nakładamy na histogram krzywą odpowiadającą rozkładowi teoretycznemu
  curve(dnorm(x, mean,sd), # rozkład teoretyczny
        add = TRUE)

  # Drugi wykres: histogram rozkładu średnich z próbek
  hist(errors,
       freq = FALSE,
       main = 'Rozkład średniej z próby',
       xlab='X') # rozkład empiryczny

  # Drugi wykres: nakładamy krzywą
  curve(dnorm(x, mean, sd/sqrt(n)),
        add = TRUE,
        col = 'red', lwd =2) # rozkład teoretyczny

  # Trzeci wykres: wykres kwantyl-kwantyl dla średnich z próbek
  qqnorm(errors, main = 'Wykres kwantyl-kwantyl')
  qqline(errors)
}
```

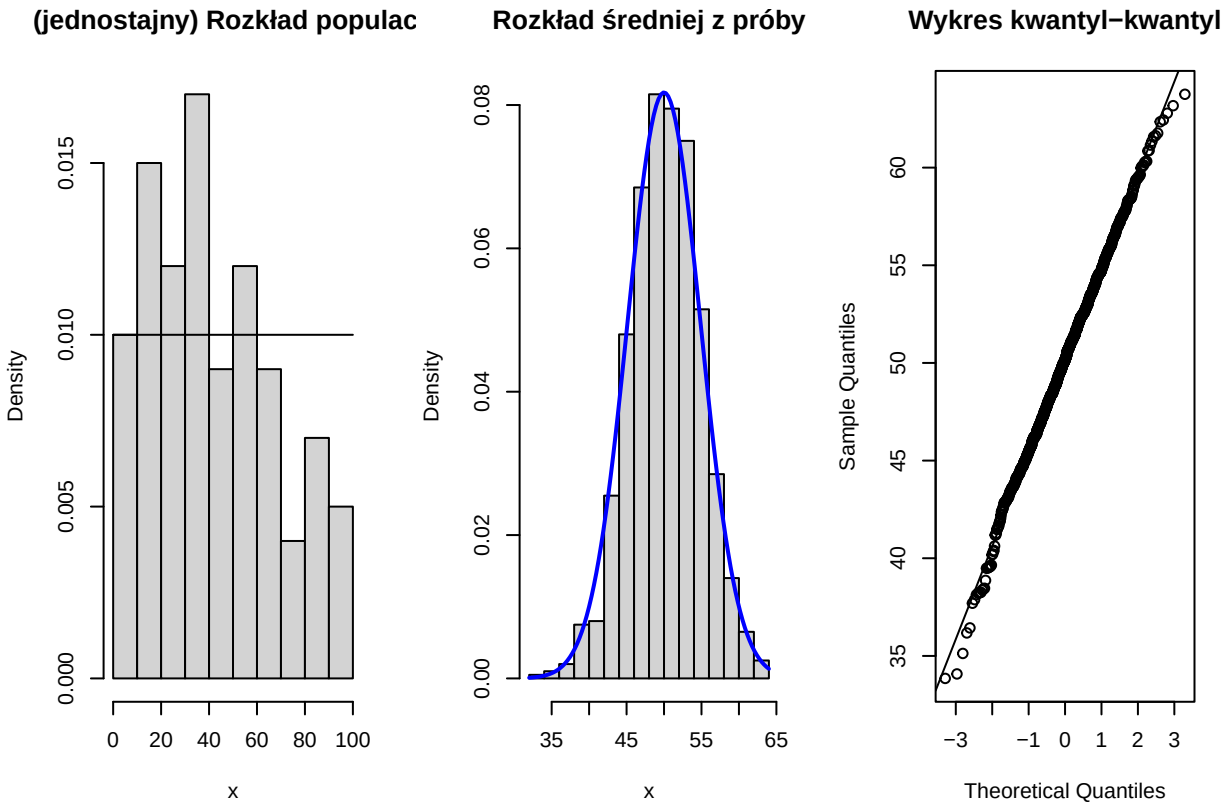
Na histogramach znajdują się "rozkłady empiryczny" to znaczy rozkłady uzyskane dzięki procedurze losowania z rozkładów o określonych parametrach. Na nie nałożone są krzywe ilustrujące gęstość prawdopodobieństwa interesujących nas rozkładów. Interesuje nas środkowy panel, na którym histogram obrazuje wyniku eksperymentu, polegającego na losowaniu prób i wyliczaniu z nich średnich, a krzywa jest krzywą rozkładu średniej z próby, o której mówi nam przywołane wcześniej sformułowanie CTG:

```
set.seed(1234)
blad_standardowy(35, i=1000)
```



Z punktu widzenia logiki testowania hipotez istotne jest, że rozkład średniej z próby będzie dążył do rozkładu normalnego nawe wówczas, gdy rozkład w populacji nie będzie normalny! Aby to zilustrować powtórzmy symulację tym razem losując nie z rozkładu normalnego, ale z rozkładu jednostajnego (*uniform distribution*, `*unif`).

```
blad_standardowy_unif <- function(n,a,b,i){
  mean <- a+b/2
  sd <- sqrt((b-a)**2/12)
  errors <- replicate(i, mean(runif(n,0,100)))
  par(mfrow=c(1,3))
  hist(runif(100,a,b),
       freq = FALSE,
       main = "(jednostajny) Rozkład populacji",
       xlab = 'x')
  curve(dunif(x, a,b), add = TRUE)
  hist(errors,
       freq = FALSE,
       main = "Rozkład średniej z próby", xlab = 'x')
  curve(dnorm(x, mean, sd/sqrt(n)),
        add = TRUE, col = 'blue', lwd=2)
  qqnorm(errors, main = 'Wykres kwantyl-kwantyl')
  qqline(errors)
}
set.seed(1234)
blad_standardowy_unif(35, 0,100, i=1000)
```



Jak widzimy na środkowym panelu rozkład średniej z próby przy $n = 35$ dąży do rozkładu normalnego!

Testowanie hipotez: przykład V

Spróbujmy teraz zastosować nasza wiedzę o Centralnym Twierdzeniu Granicznym do przetestowania hipotezy na temat średniej w populacji!

Williamson postanowił sprawdzić hipotezę, zgodnie z którą stres w życiu dziecka prowadzić ma do problemów behawioralnych. Spodziewał się, że w próbie dzieci, których rodzice mają depresję, poziom problemów behawioralnych będzie szczególnie wysoki.

Williamson przebadał 166 dzieci z domów, w których przynajmniej jeden z rodziców miał stwierdzoną historię depresji. Dzieci wypełniły *Youth Self-Report* i średnia w próbie wyniosła 55.71 z odchyleniem standardowym 7.35.

Skądinąd znamy rozkład w populacji wyników *Youth Self-Report* i wiemy, że jest nim rozkład normalny o parametrach $\mu = 50$ oraz $\sigma = 10$.

Zobaczmy jak wygląda nasz układ hipotez.

Hipoteza alternatywna (dwustronna): Te dzieci nie pochądzą z populacji o tej samej średniej YSR, co ogólna populacja.

$$H_A : \mu_1 \neq 50$$

Hipoteza alternatywna (prawostronna): Te dzieci pochodzą z populacji o wyższej średniej YSR, niż ogólna populacja.

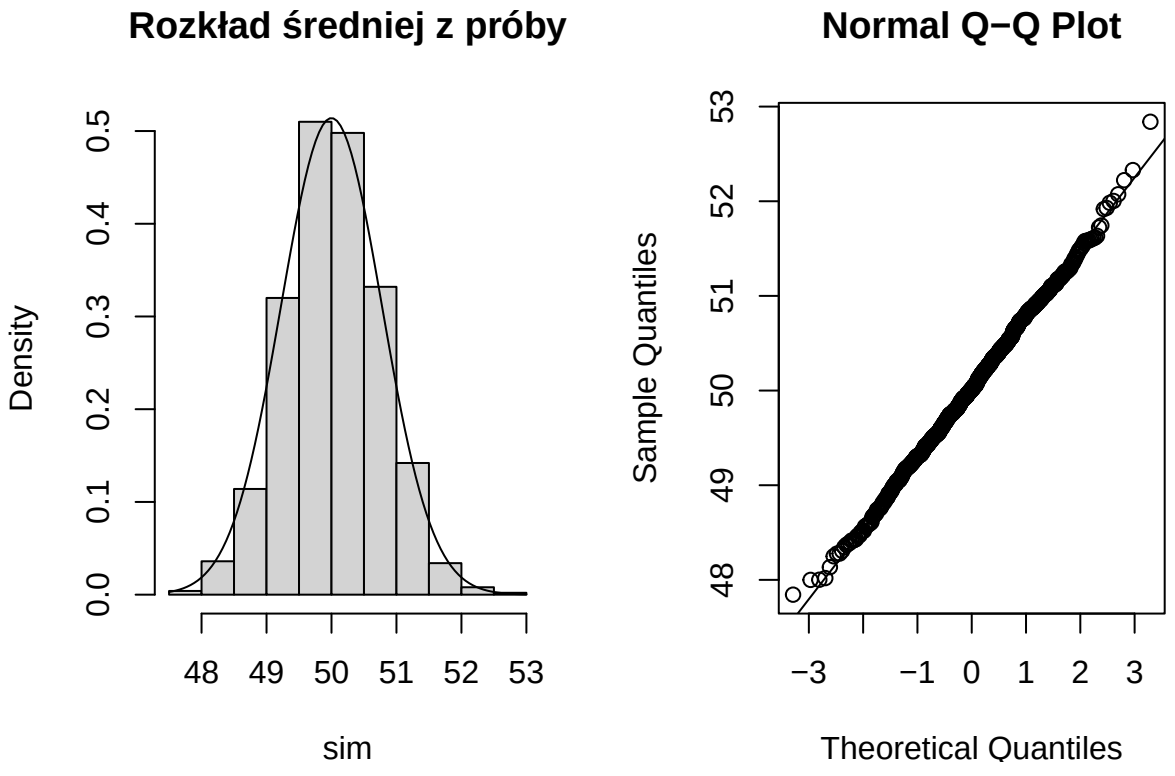
$$H_A : \mu_1 > 50$$

Hipoteza zerowa: Te dzieci pochodzą z populacji o innej średniej YSR, niż ogólna populacja.

$$H_0 : \mu_2 = 50$$

Możemy więc zobaczyć, jak wygląda rozkład średniej z próby przy założeniu prawdziwości hipotezy zerowej. W tym celu, dla ilustracji, wylosujemy z rozkładu normalnego o średniej 50 i odchyleniu standardowym 10 tysiąc 166 elementowych próbek. Z każdej z nich obliczymy średnią i za pomocą histogramu zobrazujemy ich rozkład. Dodatkowo, korzystając z Centralnego Twierdzenia Granicznego, na histogram nałożymy krzywą rozkładu normalnego o średniej 50 oraz o odchyleniu standardowym wynoszącym $\frac{\sigma}{\sqrt{n}}$, to znaczy w naszym przypadku $\frac{10}{\sqrt{166}}$. Wykres kwantyl-kwantyl posłuży nam do sprawdzenia, czy kształt tego rozkładu faktycznie zbliża się do rozkładu normalnego.

```
s_mean <- 50
n <- 166
s_sd <- 10/sqrt(n)

par(mfrow = c(1,2))
sim = replicate(1000, # losujemy 100 próbek i obliczamy średnią
                mean(rnorm(n,50,10))
                )
hist(sim,
     freq = FALSE,
     main = "Rozkład średniej z próby")
curve(dnorm(x,
            s_mean,
            s_sd),
      add = TRUE)
qqnorm(sim)
qqline(sim)
```



Teraz możemy wyznaczyć regiony odrzucenia dla hipotezy dwustronnej i prawostronnej.

Hipoteza dwustronna:

```
qnorm(c(0.025, 0.975), s_mean, s_sd)
```

```
## [1] 48.47877 51.52123
```

```
# możemy posłużyć się naszym rozkładem empirycznym jako przybliżeniem rozkładu teoretycznego!
quantile(sim, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 48.56754 51.44774
```

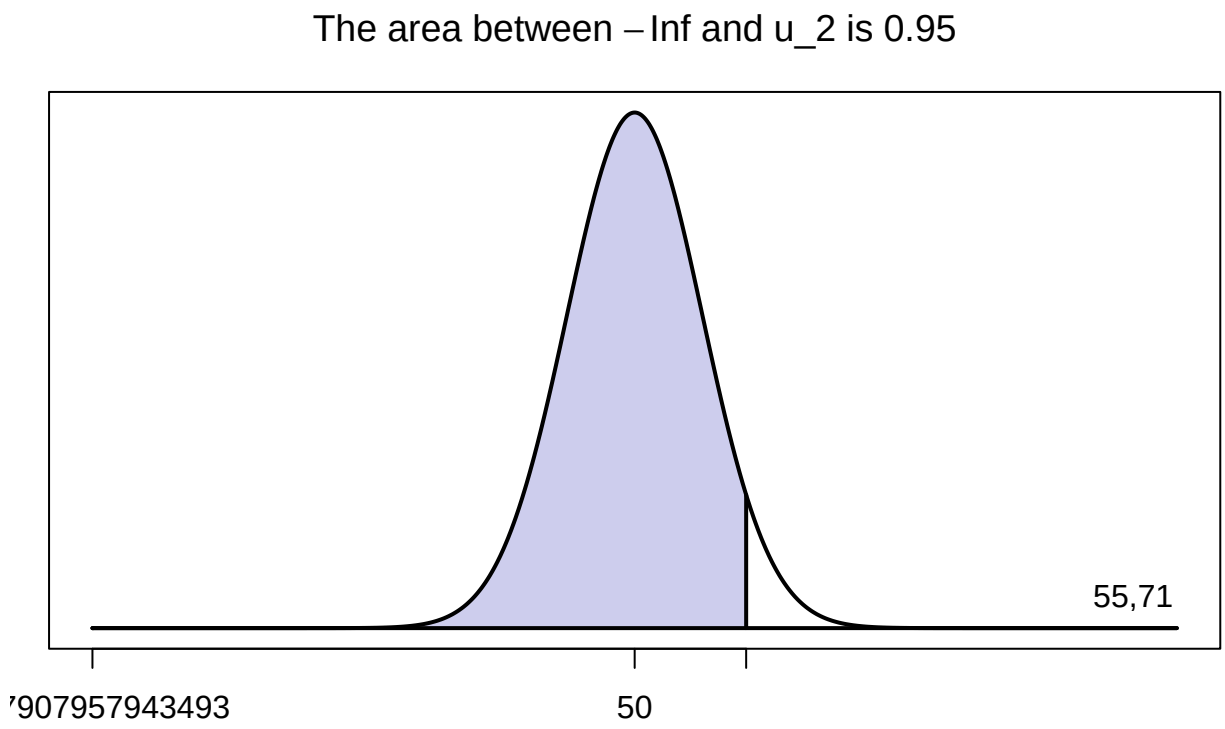
```
u_2 = qnorm(0.95, s_mean, s_sd)
print('Wartość krytyczna dla hipotezy prawostronnej')
## [1] "Wartość krytyczna dla hipotezy prawostronnej"
u_2
## [1] 51.27665
```

```
l = qnorm(0.025, s_mean, s_sd)
u = qnorm(0.975, s_mean, s_sd)
print('Wartości krytyczne dla hipotezy dwustronnej')
## [1] "Wartości krytyczne dla hipotezy dwustronnej"
l
## [1] 48.47877
u
## [1] 51.52123
```

Odpowiednie obszary odrzucenia możemy również nanieść na wykres.

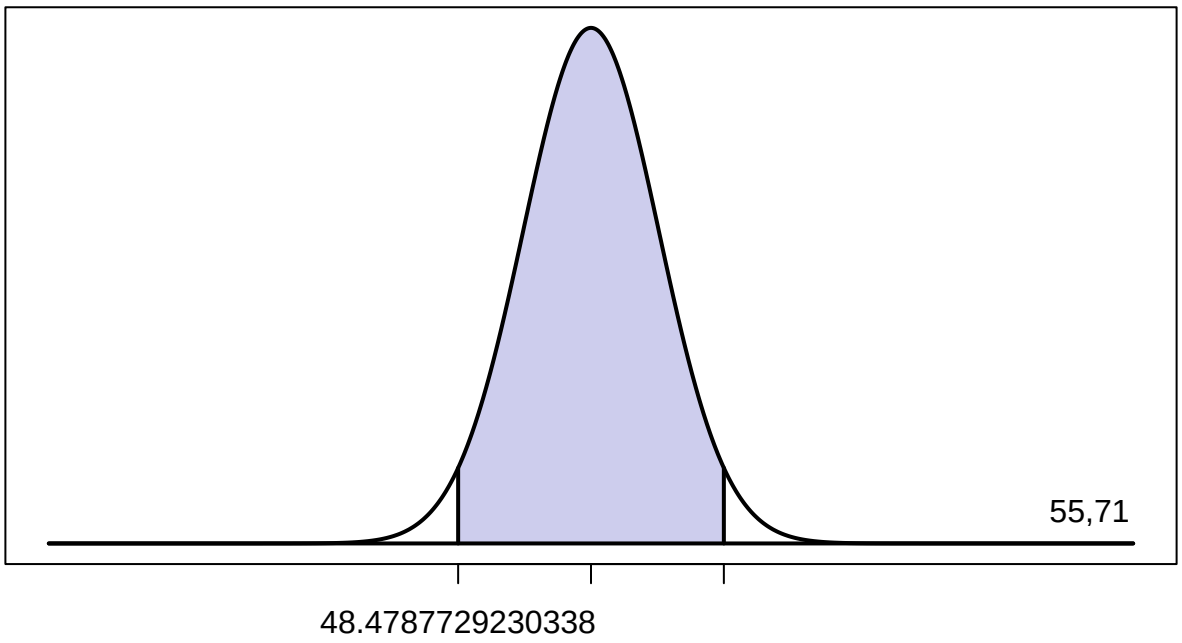
```

norm_area(-Inf, u_2,  s_mean, s_sd)
text(labels = "55,71", x = 55.71, y = 0.03)
```

Obszar krytyczny dla hipotezy prawostronnej (biały kolor)




```
norm_area(l, u, s_mean, s_sd)
text(labels = "55,71", x = 55.71, y = 0.03)
```

Obszary krytyczne dla hipotezy niekierunkowej (biały kolor)



Jak widzimy dla obu rodzajów testów faktycznie uzyskana przez nas średnia z próby (55.71) wpada w obszar odrzucenia. Korzystając z dystrybuanty rozkładu normalnego, możemy obliczyć dokładne wartości p dla obu rodzajów testów.

```
print('Prawdopodobieństwo uzyskania średniej z próby 55.71\
      przy założeniu hipotezy zerowej dla hipotezy prawostronnej')
## [1] "Prawdopodobieństwo uzyskania średniej z próby 55.71\n      przy założeniu hipotezy zerowej dla h
pnorm(55.71, s_mean, s_sd, lower.tail = FALSE)
## [1] 9.417126e-14
```

```
print('Prawdopodobieństwo uzyskania średniej z próby 55.71\
      przy założeniu hipotezy zerowej dla hipotezy dwustronnej')
## [1] "Prawdopodobieństwo uzyskania średniej z próby 55.71\n      przy założeniu hipotezy zerowej dla h
pnorm(55.71, s_mean, s_sd, lower.tail = FALSE) * 2
## [1] 1.883425e-13
```

Jak widzimy możemy odrzucić hipotezę zerową. Wartości p znajdują się dużo poniżej progu $\alpha = 0.05$. Uzyskanie takiej średniej jest skrajnie nieprawdopodobne. Wydaje się więc, że Williamson wykazał prawdziwość swojej tezy.

W ogólności takie postępowanie możemy nazwać **testem z**:

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

- $\bar{X}$ - średnia z próby
- μ - średnia z populacji zakładana w hipotezie zerowej
- σ - znane odchylenie standardowe w populacji
- n - liczebność próby

Tak obliczoną statystykę testową z porównujemy do standardowego rozkładu normalnego, to znaczy rozkładu

normalnego o średniej wynoszącej 0 i odchyleniu równym 1. W przypadku analizowanego przez nas badania, wyglądałoby to tak:

```
z_stat <- (55.71 - 50) / (10 / sqrt(166))
pnorm(z_stat, lower.tail = F) * 2
```

```
## [1] 1.883425e-13
```

Aproksymacja rozkładu dwumianowego z rozkładu normalnego

Rozkład normalny jest niezłym przybliżeniem rozkładu dwumianowego. Możemy się o tym przekonać wykorzystując następującą wiedzę:

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

gdzie

- n - liczba prób
- p - prawdopodobieństwo sukcesu

```
# Wylosujmy 1000 wartosci z B(100,0.5)

hist(rbinom(10000, 100, 0.5), main = 'B(100,0.5)', freq = FALSE)

# Nałóżmy krzywą N(np, sqrt(np1-p))
curve(dnorm(x, 100*0.5, sqrt(100*0.5*(1-0.5))), add = TRUE)
```