Zadanie 2. W tym zadaniu zajmować będziemy się wyznaczaniem przedziału ufności dla średniej μ w modelu normalnym o **znanej** wariancji σ^2 na poziomie ufności $1-\alpha$ dla $\alpha = 0.05$. W tym celu wykorzystywać będziemy następujący wzór na przedział ufności:

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right],$$

gdzie $z_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ standardowego rozkładu normalnego $N(0,1)$.

```
confint <- function(x, alpha, var)
return(c(mean(x) - qnorm(1 - alpha/2) * sqrt(var)/sqrt(length(x)),
         mean(x) + qnorm(1 - alpha/2) * sqrt(var)/sqrt(length(x))))
```

Zauważmy, że powyższego wzoru nie możemy zastosować w przypadku rozkładu Cauchy'ego, dla którego wartość wariancji jest niezdefiniowana.

W wykonywanym eksperymencie będziemy dla kolejno omawianych rozkładów generować 50 obserwacji, które umożliwią nam wyznaczenie przedziału ufności dla średniej μ. Następnie będziemy powtarzać to doświadczenie 10 000 razy, aby móc oszacować prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności.

```
prob <- function(distr, mean, loc, scale, var, alpha, k) {
  n <- 0
  for (i in c(1:10000)) {
    conf_int <- confint(distr(k, loc, scale), alpha, var)
    if (conf_int[1] <= mean & conf_int[2] >= mean) n <- n + 1
  }
  return(round(n/10000, 3))
}
```

W pierwszej części zajmiemy się rozkładem normalnym o parametrach kolejno $N(0,1), N(0,2)$, $N(0,3)$. Zaczniemy od wyznaczenia przedziałów ufności na podstawie 50 obserwacji wygenerowanych kolejno z każdego z powyższych rozkładów.

```
result <- round(data.frame(confint(rnorm(50, 0, 1), .05, 1),
                           confint(rnorm(50, 0, 2), .05, 4),
                           confint(rnorm(50, 0, 3), .05, 9)), 3)
```

	$N(0,1)$	$N(0,2)$	$N(0,3)$
dolna granica	-0.217	-0.232	-1.350
górna granica	0.337	0.876	0.313

Możemy zauważyć, że wraz ze wzrostem wariancji, zwiększa się szerokość przedziału. Zauważmy, że wynika to natychmiast z postaci używanego przez nas wzoru na przedział ufności. Ponadto widzimy, że szerokość przedziałów jest dość duża, a zatem używana przez nas metoda estymacji nie jest całkowicie optymalna. Powtarzamy powyższy eksperyment 10 000 razy i wyznaczamy prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności.

	$N(0,1)$	$N(0,2)$	$N(0,3)$
prawdopodobieństwo	0.95	0.95	0.95

Widzimy zatem, że prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności w każdym przypadku jest zbliżone do 0.95, tj. wartości $1-\alpha$.

Powtarzamy powyższy eksperyment dla rozkładu logistycznego o parametrach kolejno $L(0,1)$, $L(0,2), L(0,3)$. Wariancje oraz średnie rozkładów są wówczas równe:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
średnia	0.00	0.000	0.000
wariancja	3.29	13.159	29.609

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
dolna granica	-0.399	-1.043	-1.068
górna granica	0.607	0.968	1.949

Zauważmy, że powstałe przedziały są szersze od tych dla rozkładu normalnego. W tym przypadku prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności znów zbliżone jest do 0.95:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
prawdopodobieństwo	0.95	0.954	0.951

Przejdziemy teraz do powtórzenia powyższego eksperymentu dla rozkładu wykładniczego o parametrach kolejno $\exp(1), \exp(1/2), \exp(1/3)$. Wariancje oraz średnie rozkładów są wówczas równe:

	exp(1)	exp(1/2)	exp(1/3)
średnia	1	2	3
wariancja	1	4	9

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	exp(1)	exp(1/2)	exp(1/3)
dolna granica	0.759	1.307	2.732
górna granica	1.313	2.415	4.395

Zauważmy, że powstałe przedziały są znów znacznie szersze od tych dla rozkładu normalnego. Następnie używając zmodyfikowanej funkcji `prob` szacujemy prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności. Znów okaże się być ono bliskie 0.95.

	exp(1)	exp(1/2)	exp(1/3)
prawdopodobieństwo	0.947	0.955	0.951

Na koniec powtórzymy powyższy eksperyment dla rozkładu chi-kwadrat kolejno $\chi^2(1), \chi^2(2), \chi^2(3)$. Wariancje oraz średnie rozkładów są wówczas równe:

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
średnia	1	2	3
wariancja	2	4	6

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
dolna granica	0.381	0.735	2.131
górna granica	1.165	1.843	3.489

Zgodnie z przewidywaniami powstałe przedziały są znów znacznie szersze od tych dla rozkładu normalnego. Podobnie jak wcześniej oszacowane prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności w każdym przypadku jest zbliżone do 0.95.

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
prawdopodobieństwo	0.952	0.952	0.95

Zadanie 4. W tym zadaniu zajmować będziemy się wyznaczaniem przedziału ufności dla średniej μ w modelu normalnym o **nieznanej** wariancji σ^2 na poziomie ufności $1-\alpha$ dla $\alpha = 0.05$. W tym celu wykorzystywać będziemy następujący wzór na przedział ufności:

$$\left[\bar{X} - t_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}},\ \bar{X} + t_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\right],$$

gdzie $t_{(1-\frac{\alpha}{2},n-1)}$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ rozkładu t-Studenta z $n-1$ stopniami swobody.

```
confint <- function(x, alpha)
return(c(mean(x) - qt(1 - alpha/2, length(x) - 1) * sd(x)/sqrt(length(x)),
         mean(x) + qt(1 - alpha/2, length(x) - 1) * sd(x)/sqrt(length(x))))
```

W wykonywanym eksperymencie będziemy dla kolejno omawianych rozkładów generować 50 obserwacji, które umożliwią nam wyznaczenie przedziału ufności dla średniej μ. Następnie będziemy powtarzać to doświadczenie 10 000 razy, aby móc oszacować prawdopodobieństwo pokryci nieznanej średniej przez przedział ufności.

```
prob <- function(distr, mean, loc, scale, alpha, k) {
  n <- 0
  for (i in c(1:10000)) {
    conf_int <- confint(distr(k, loc, scale), alpha)
    if (conf_int[1] <= mean & conf_int[2] >= mean) n <- n + 1
  }
  return(round(n/10000, 3))
}
```

W pierwszej części zajmiemy się rozkładem normalnym o parametrach kolejno $N(0,1), N(0,2)$, $N(0,3)$. Przedziały ufności na podstawie 50 obserwacji wygenerowanych kolejno z każdego z powyższych rozkładów wynoszą:

```
result <- round(data.frame(confint(rnorm(50, 0, 1), .05),
                           confint(rnorm(50, 0, 2), .05),
                           confint(rnorm(50, 0, 3), .05)), 3)
```

	$N(0,1)$	$N(0,2)$	$N(0,3)$
dolna granica	-0.362	-0.173	-1.494
górna granica	0.169	1.048	0.302

Możemy zauważyć, że znów wraz ze wzrostem wariancji, zwiększa się szerokość przedziału. Ponadto widzimy, że szerokość przedziałów jest dość szeroka. Powtarzamy powyższy eksperyment 10 000 razy i wyznaczamy prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności.

	$N(0,1)$	$N(0,2)$	$N(0,3)$
prawdopodobieństwo	0.945	0.949	0.948

Widzimy zatem, że prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności w każdym przypadku jest zbliżone do 0.95, tj. wartości $1-\alpha$.

Powtarzamy powyższy eksperyment dla rozkładu logistycznego o parametrach kolejno $L(0,1)$, $L(0,2)$, $L(0,3)$.

Dla losowo wygenerowanych 50 obserwacji z w.w. rozkładów otrzymujemy następujące przedziały ufności:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
dolna granica	-0.636	-0.706	-2.117
górna granica	0.466	1.329	1.001

Zauważmy, że powstałe przedziały są szersze od tych dla rozkładu normalnego. W tym przypadku prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności znów jest bliskie 0.95.

	$L(0,1)$	$L(0,2)$	$L(0,3)$
prawdopodobieństwo	0.948	0.952	0.95

Następnie powtórzymy pierwszą część powyższego doświadczenia dla rozkładów Cauchy'ego o parametrach kolejno $C(0,1)$, $C(0,2)$, $C(0,3)$. Nie będziemy jednak badać prawdopodobieństwa pokrycia wartości średniej przez wygenerowany przedział, ponieważ z definicji wartość średnia dla rozkładu Cauchy'ego jest niezdefiniowana.

	$C(0,1)$	$C(0,2)$	$C(0,3)$
dolna granica	-13.730	-2.030	-0.962
górna granica	5.201	2.144	6.629

Możemy zauważyć, że przedziały są bardzo szerokie i nie ma w nich żadnej zależności. Przejdziemy teraz do powtórzenia powyższego eksperymentu dla rozkładu wykładniczego kolejno $\exp(1)$, $\exp(1/2)$, $\exp(1/3)$. Generując 50 obserwacji z w.w. rozkładów otrzymujemy następujące przedziały ufności:

	exp(1)	exp(1/2)	exp(1/3)
dolna granica	0.840	1.558	2.569
górna granica	1.565	2.708	4.512

Zauważmy, że powstałe przedziały są znów znacznie szersze od tych dla rozkładu normalnego. Następnie używając zmodyfikowanej funkcji `prob` szacujemy prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności.

	exp(1)	exp(1/2)	exp(1/3)
prawdopodobieństwo	0.934	0.937	0.936

Zauważmy, że w tym przypadku, inaczej niż w poprzednich doświadczeniach, prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności w każdym przypadku jest zbliżone do **0.93**.

Na koniec powtórzymy powyższy eksperyment dla rozkładu chi-kwadrat kolejno $\chi^2(1), \chi^2(2), \chi^2(3)$. Otrzymujemy następujące przedziały ufności:

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
dolna granica	0.591	1.280	2.104
górna granica	1.185	2.155	3.532

Powstałe przedziały są znów znacznie szersze od tych dla rozkładu normalnego. Ponadto oszacowane prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności znów wynosi w przybliżeniu **0.93**.

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
prawdopodobieństwo	0.924	0.936	0.935

Zadanie 6. W tym zadaniu zajmować będziemy się wyznaczaniem przedziału ufności dla wariancji σ^2 w modelu normalnym o **znanej** średniej μ na poziomie ufności $1-\alpha$ dla $\alpha = 0.05$. W tym celu wykorzystywać będziemy następujący wzór na przedział ufności:

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_{1-\frac{\alpha}{2},n-1}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_{\frac{\alpha}{2},n-1}}\right],$$

gdzie $\chi_{1-\frac{\alpha}{2},n-1}, \chi_{\frac{\alpha}{2},n-1}$ są kwantylami rzędów kolejno $1-\frac{\alpha}{2}, \frac{\alpha}{2}$ z rozkładu χ^2 z $n-1$ stopniami swobody, zaś n jest wielkością próby. Zauważmy, że podobnie jak w zadaniu 1., powyższy wzór nie dostarczy nam informacji n.t. przedziału ufności dla wariancji dla rozkładu Cauchy'ego, dlatego że w tym rozkładzie średnia jest niezdefiniowana.

```
confint <- function(x, alpha, mean)
  return(c(sum((x - mean)^2)/qchisq(1 - alpha/2, length(x) - 1),
           sum((x - mean)^2)/qchisq(alpha/2, length(x) - 1)))
```

W wykonywanym eksperymencie będziemy dla kolejno omawianych rozkładów generować 50 obserwacji, które umożliwią nam wyznaczenie przedziału ufności dla wariancji σ^2. Następnie będziemy powtarzać to doświadczenie 10 000 razy, aby móc oszacować prawdopodobieństwo pokryci nieznanej wariancji przez przedział ufności.

```
prob <- function(distr, mean, loc, scale, var, alpha, k) {
  n <- 0
  for (i in c(1:10000)) {
    conf_int <- confint(distr(k, loc, scale), alpha, mean)
    if (conf_int[1] <= var & conf_int[2] >= var) n <- n + 1
  }
  return(round(n/10000, 3))
}
```

W pierwszej części zajmiemy się rozkładem normalnym o parametrach kolejno $N(0,1), N(0,2)$, $N(0,3)$. Zaczniemy od wyznaczenia przedziałów ufności na podstawie 50 obserwacji wygenerowanych kolejno z każdego z powyższych rozkładów.

```
result <- round(data.frame(confint(rnorm(50, 0, 1), .05, 0),
                           confint(rnorm(50, 0, 2), .05, 0),
                           confint(rnorm(50, 0, 3), .05, 0)), 3)
```

	$N(0,1)$	$N(0,2)$	$N(0,3)$
dolna granica	0.854	3.229	4.395
górna granica	1.900	7.185	9.781

Widzimy, że szerokość przedziałów jest bardzo duża, co świadczy o nieoptymalności dobranej metody estymacji. Powtarzamy powyższy eksperyment 10 000 razy i wyznaczamy prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności.

	$N(0,1)$	$N(0,2)$	$N(0,3)$
prawdopodobieństwo	0.95	0.945	0.952

Widzimy zatem, że prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności w każdym przypadku jest zbliżone do 0.95, tj. wartości $1-\alpha$.

Powtarzamy powyższy eksperyment dla rozkładu logistycznego o parametrach kolejno $L(0,1)$, $L(0,2)$, $L(0,3)$. Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
dolna granica	2.420	12.121	20.996
górna granica	5.386	26.974	46.724

Zauważmy, że powstałe przedziały są szersze od tych dla rozkładu normalnego. W tym przypadku prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności wynosi jest zbliżone do 0.88, co stanowi dość słaby rezultat.

	$L(0,1)$	$L(0,2)$	$L(0,3)$
prawdopodobieństwo	0.891	0.887	0.885

Przejdziemy teraz do powtórzenia powyższego eksperymentu dla rozkładu wykładniczego kolejno $\exp(1), \exp(1/2), \exp(1/3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	exp(1)	exp(1/2)	exp(1/3)
dolna granica	1.395	3.662	4.322
górna granica	3.104	8.150	9.617

Następnie używając zmodyfikowanej funkcji prob szacujemy prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności.

	exp(1)	exp(1/2)	exp(1/3)
prawdopodobieństwo	0.714	0.714	0.71

Widzimy, że dla rozkładu wykładniczego prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności jest jeszcze niższe niż w przypadku rozkładu logistycznego. Wynosi w przybliżeniu jedynie 0.70. Stąd jako wniosek otrzymujemy, że wykonywana metoda estymacji wariancji nie jest tutaj skuteczna.

Na koniec powtórzymy powyższy eksperyment dla rozkładu chi-kwadrat kolejno $\chi^2(1), \chi^2(2), \chi^2(3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
dolna granica	1.298	3.255	3.360
górna granica	2.888	7.244	7.478

Następnie, szacując prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności, widzimy, że dla rozkładu chi-kwadrat, podobnie jak w przypadku rozkładu wykładniczego, oszacowanie wariancji jest znacznie poniżej oczekiwanego poziomu.

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
prawdopodobieństwo	0.59	0.714	0.783

Zadanie 8. W tym zadaniu zajmować będziemy się wyznaczaniem przedziału ufności dla wariancji σ^2 w modelu normalnym o **nieznanej** średniej μ na poziomie ufności $1-\alpha$ dla $\alpha = 0.05$. W tym celu wykorzystywać będziemy następujący wzór na przedział ufności:

$$\left[\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2},n-1}}\right],$$

gdzie $\chi_{1-\frac{\alpha}{2},n-1}, \chi_{\frac{\alpha}{2},n-1}$ są kwantylami rzędów kolejno $1-\frac{\alpha}{2}, \frac{\alpha}{2}$ z rozkładu χ^2 z $n-1$ stopniami swobody, zaś n jest wielkością próby.

```
confint <- function(x, alpha, mean)
  return(c(((length(x)-1)*var(x))/qchisq(1 - alpha/2, length(x) - 1),
           ((length(x)-1)*var(x))/qchisq(alpha/2, length(x) - 1)))
```

W wykonywanym eksperymencie będziemy dla kolejno omawianych rozkładów generować 50 obserwacji, które umożliwią nam wyznaczenie przedziału ufności dla wariancji σ^2. Następnie będziemy powtarzać to doświadczenie 10 000 razy, aby móc oszacować prawdopodobieństwo pokryci nieznanej wariancji przez przedział ufności.

```
prob <- function(distr, mean, loc, scale, var, alpha, k) {
  n <- 0
  for (i in c(1:10000)) {
    conf_int <- confint(distr(k, loc, scale), alpha, mean)
    if (conf_int[1] <= var & conf_int[2] >= var) n <- n + 1
  }
  return(round(n/10000, 3))
}
```

W pierwszej części zajmiemy się rozkładem normalnym o parametrach kolejno $N(0,1), N(0,2)$, $N(0,3)$. Zaczniemy od wyznaczenia przedziałów ufności na podstawie 50 obserwacji wygenerowanych kolejno z każdego z powyższych rozkładów.

```
result <- round(data.frame(confint(rnorm(50, 0, 1), .05, 0),
                           confint(rnorm(50, 0, 2), .05, 0),
                           confint(rnorm(50, 0, 3), .05, 0)), 3)
```

	$N(0,1)$	$N(0,2)$	$N(0,3)$
dolna granica	0.879	3.305	9.017
górna granica	1.955	7.355	20.066

Widzimy, że podobnie jak w przypadku znanej średeniej, szerokość przedziałów jest bardzo szeroka, co świadczy o źle dobranej metodzie estymacji. Powtarzamy powyższy eksperyment 10 000 razy i wyznaczamy prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności.

	$N(0,1)$	$N(0,2)$	$N(0,3)$
prawdopodobieństwo	0.947	0.952	0.945

Widzimy zatem, że prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności w każdym przypadku jest zbliżone do 0.95.

Powtarzamy powyższy eksperyment dla rozkładu logistycznego o parametrach kolejno $L(0,1)$, $L(0,2)$, $L(0,3)$.

Generując 50 obserwacji z w.w. rozkładów otrzymujemy następujące przedziały ufności:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
dolna granica	2.441	8.713	27.999
górna granica	5.432	19.390	62.310

Zauważmy, że powstałe przedziały są szersze od tych dla rozkładu normalnego. W tym przypadku prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności wynosi:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
prawdopodobieństwo	0.882	0.884	0.891

Widzimy, że prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności w każdym przypadku jest zbliżone do 0.88, co stanowi dość słaby rezultat.

Następnie powtórzymy pierwszą część powyższego doświadczenia dla rozkładów Cauchy'ego o parametrach kolejno $C(0,1)$, $C(0,2)$, $C(0,3)$. Nie będziemy jednak badać prawdopodobieństwa pokrycia wartości wariancji przez wygenerowany przedział, ponieważ z definicji wariancja dla rozkładu Cauchy'ego jest niezdefiniowana.

	$C(0,1)$	$C(0,2)$	$C(0,3)$
dolna granica	11.290	194.142	66.263
górna granica	25.124	432.044	147.462

Możemy zauważyć, że przedziały są bardzo szerokie i nie ma w nich żadnej zależności.

Przejdziemy teraz do powtórzenia powyższego eksperymentu dla rozkładu wykładniczego o parametrach kolejno $\exp(1), \exp(1/2), \exp(1/3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	exp(1)	exp(1/2)	exp(1/3)
dolna granica	0.708	3.449	5.332
górna granica	1.577	7.674	11.865

Następnie używając zmodyfikowanej funkcji prob szacujemy prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności.

	exp(1)	exp(1/2)	exp(1/3)
prawdopodobieństwo	0.704	0.702	0.704

Widzimy, że dla rozkładu wykładniczego prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności jest jeszcze niższe niż w przypadku rozkładu logistycznego. Wynosi w przybliżeniu jedynie 0.70. Stąd jako wniosek otrzymujemy, że wykonywana metoda estymacji wariancji nie jest tutaj skuteczna.

Na koniec powtórzymy powyższy eksperyment dla rozkładu chi-kwadrat o stopniach swobody równych kolejno $\chi^2(1), \chi^2(2), \chi^2(3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
dolna granica	0.688	3.198	2.505
górna granica	1.530	7.117	5.574

Następnie szacujemy prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności.

	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$
prawdopodobieństwo	0.578	0.706	0.773

Widzimy więc, że dla rozkładu chi-kwadrat, podobnie jak w przypadku rozkładu wykładniczego, oszacowanie wariancji jest znacznie poniżej oczekiwanego poziomu.

Zauważmy, że wnioski otrzymane dla przedziału ufności dla wariancji przy nieznanej średniej są analogiczne do tych otrzymanych w zadaniu 6. Widzimy zatem, że nie powinniśmy używać omawianych metod estymacji wariancji właściwych dla rozkładu normalnego do innych rozkładów.

Zadanie 8.

W tym zadaniu zajmować będziemy się wyznaczaniem przedziału ufności dla proporcji p w modelu normalnym na poziomie ufności $1-\alpha$ dla $\alpha = 0.05$. W tym celu wykorzystywać będziemy następujący wzór:

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right],$$

gdzie $z_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ standardowego rozkładu normalnego $N(0,1)$., zaś n jest wielkością próby.

```
confint <- function(x, alpha)
  return(c(mean(x)-qnorm(1-alpha/2)*sqrt(abs(mean(x)*(1-mean(x)))/length(x)),
           mean(x)+qnorm(1-alpha/2)*sqrt(abs(mean(x)*(1-mean(x)))/length(x))))
```

W wykonywanym eksperymencie będziemy dla kolejno omawianych rozkładów generować 50 obserwacji, które umożliwią nam wyznaczenie przedziału ufności dla proporcji. Następnie będziemy powtarzać to doświadczenie 10 000 razy, aby móc oszacować prawdopodobieństwo pokrycia nieznanej proporcji przez przedział ufności.

```
prob <- function(distr, mean, sigma, prop, alpha, k) {
  n <- 0
  for (i in c(1:10000)) {
    conf_int <- confint(distr(k, mean, sigma), alpha)
    if (conf_int[1] <= prop & conf_int[2] >= prop) n <- n + 1
  }
  return(n/10000)
}
```

W pierwszej części zajmiemy się rozkładem normalnym o parametrach kolejno $N(0,1), N(0,2)$, $N(0,3)$. Zaczniemy od wyznaczenia przedziałów ufności na podstawie 50 obserwacji wygenerowanych kolejno z każdego z powyższych rozkładów.

```
result <- round(data.frame(confint(rnorm(50, 0, 1), .05),
                           confint(rnorm(50, 0, 2), .05),
                           confint(rnorm(50, 0, 3), .05)), 3)
```

	$N(0,1)$	$N(0,2)$	$N(0,3)$
dolna granica	-0.019	0.064	0.557
górna granica	0.054	0.271	0.815

Powtarzamy powyższy eksperyment 10 000 razy i wyznaczamy prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności.

	$N(0,1)$	$N(0,2)$	$N(0,3)$
prawdopodobieństwo	0.004	0.086	0.129

Widzimy zatem, że jest ono znikome i rośnie wraz ze wzrostem wariancji.

Powtarzamy powyższy eksperyment dla rozkładu logistycznego o parametrach kolejno $L(0,1)$, $L(0,2), L(0,3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
dolna granica	0.272	-0.085	0.980
górna granica	0.545	0.018	1.059

W tym przypadku prawdopodobieństwo pokrycia nieznanej proporcji przez przedział ufności wynosi:

	$L(0,1)$	$L(0,2)$	$L(0,3)$
prawdopodobieństwo	0.07	0.07	0.073

Widzimy, że znów prawdopodobieństwo to jest nieznaczne.

Powtarzamy powyższy eksperyment dla rozkładu Cauchy'ego o parametrach kolejno $C(0,1)$, $C(0,2), C(0,3)$.

Podobnie jak wcześniej generujemy 50 obserwacji z w.w. rozkładów i otrzymujemy następujące przedziały ufności:

	$C(0,1)$	$C(0,2)$	$C(0,3)$
dolna granica	0.725	0.111	2.257
górna granica	0.933	0.343	3.564

Zauważmy, że powstałe przedziały, podobnie ja w przypadku rozkładu normalnego, są dosyć szerokie. W tym przypadku prawdopodobieństwo pokrycia nieznanej proporcji przez przedział ufności wynosi:

	$C(0,1)$	$C(0,2)$	$C(0,3)$
prawdopodobieństwo	0.062	0.069	0.065

Podobnie jak w dla poprzednich rozkładów prawdopodobieństwo to jest niskie. Jako wniosek otrzymujemy więc przypuszczenie, że używana przez nas metoda estymacji proporcji dodatnich obserwacji jest nieskuteczna dla rozkładów ciągłych.

Zadanie 11. W tym zadaniu powtórzymy omawiane wcześniej eksperymenty numeryczne dla prób 20-elementowych oraz 100-elementowych i rozkładów kolejno $N(0,1)$, $L(0,1)$, $C(0,1)$, $\exp(1)$, $\chi^2(1)$.

Dla rozkładu $N(0,1)$ otrzymujemy następujące przedziały ufności dla średniej i **znanej** wariancji:

	20 obserwacji	100 obserwacji
dolna granica	-0.668	-0.180
górna granica	0.209	0.212

zaś prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności wynosi:

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.953	0.951

Przejdzimy teraz do estymacji średniej dla **nieznanej** wariancji. Dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	-0.283	-0.168
górna granica	0.854	0.198

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.95	0.949

Przejdzimy teraz do estymacji wariancji. Dla **znanej** średniej dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	1.072	0.878
górna granica	3.954	1.537

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.951	0.948

Dla **nieznanej** średniej estymując wariancję otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	0.647	0.757
górna granica	2.388	1.325

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.951	0.947

Widzimy, że zgodnie z przewidywaniami, dla większej próby uzyskiwane przedziały są dokładniejsze.

Na koniec powtórzymy eksperyment z zadania 10. kolejno dla 20 obserwacji i dla 100 obserwacji. Otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	-0.298	-0.122
górna granica	0.038	-0.016

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.0901	0

Widzimy, że wraz ze zwiększeniem liczebności próby zmniejsza się prawdopodobieństwo pokrycia proporcji dodatnich obserwacji przez wygenerowany przedział ufności.

Dla rozkładu $L(0, 1)$ otrzymujemy następujące przedziały ufności dla średniej i **znanej** wariancji:

	20 obserwacji	100 obserwacji
dolna granica	-0.784	-0.214
górna granica	0.805	0.497

zaś prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności wynosi:

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.946	0.947

Przejdzimy teraz do estymacji średniej dla **nieznanej** wariancji. Dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	-1.545	-0.284
górna granica	0.543	0.421

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.951	0.951

Przejdzimy teraz do estymacji wariancji. Dla **znanej** średniej dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	2.017	2.167
górna granica	7.441	3.793

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.9	0.885

Dla **nieznanej** średniej estymując wariancję otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	1.649	1.868
górna granica	6.081	3.270

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.897	0.886

Widzimy, że zgodnie z przewidywaniami, dla większej próby uzyskiwane przedziały są dokładniejsze.

Na koniec powtórzymy eksperyment z zadania 10. kolejno dla 20 obserwacji i dla 100 obserwacji. Otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	-0.133	-0.252
górna granica	0.050	-0.080

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.184	0.0126

Widzimy, że podobnie jak dla rozkładu normalnego, wraz ze zwiększeniem liczebności próby zmniejsza się prawdopodobieństwo pokrycia proporcji dodatnich obserwacji przez wygenerowany przedział ufności.

Dla rozkładu $C(0,1)$ otrzymujemy następujące wyniki dla średniej i **nieznanej** wariancji:

	20 obserwacji	100 obserwacji
dolna granica	-4.608	-1.895
górna granica	16.334	0.986

Przejdzimy teraz do estymacji wariancji. Dla **nieznanej** średniej estymując wariancję otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	2.338	12.171
górna granica	8.623	21.306

Widzimy, że zgodnie z przewidywaniami, dla większej próby uzyskiwane przedziały są dokładniejsze.

Na koniec powtórzymy eksperyment z zadania 10. kolejno dla 20 obserwacji i dla 100 obserwacji. Otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	1.044	-2.856
górna granica	1.635	-1.771

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.1019	0.051

Widzimy, że wyniki są analogiczne do tych uzyskanych wcześniej. Przejdziemy teraz do rozkładu $\exp(1)$. Na początek prezentujemy przedziały ufności dla średniej i **znanej** wariancji:

	20 obserwacji	100 obserwacji
dolna granica	0.384	0.900
górna granica	1.261	1.292

zaś prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności wynosi:

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.952	0.955

Przejdzimy teraz do estymacji średniej dla **nieznanej** wariancji. Dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	0.654	0.739
górna granica	1.867	1.032

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.918	0.944

Przejdzimy teraz do estymacji wariancji. Dla **znanej** średniej dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	1.095	0.441
górna granica	4.039	0.771

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.772	0.688

Dla **nieznanej** średniej estymując wariancję otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	0.42	0.561
górna granica	1.55	0.982

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.728	0.685

Widzimy, że zgodnie z przewidywaniami, dla większej próby uzyskiwane przedziały są dokładniejsze. Poza tym uzyskane wnioski są podobne do tych, uzyskanych dla próby 50 elementowej.

Przejdziemy teraz do rozkładu $\chi^2(1)$. Na początek prezentujemy przedziały ufności dla średniej i **znanej** wariancji:

	20 obserwacji	100 obserwacji
dolna granica	0.651	0.769
górna granica	1.891	1.323

zaś prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności wynosi:

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.954	0.95

Przejdzimy teraz do estymacji średniej dla **nieznanej** wariancji. Dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	0.545	0.619
górna granica	2.225	1.328

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.888	0.937

Przejdzimy teraz do estymacji wariancji. Dla **znanej** średniej dostajemy:

	20 obserwacji	100 obserwacji
dolna granica	0.802	1.341
górna granica	2.957	2.347

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.622	0.564

Dla **nieznanej** średniej estymując wariancję otrzymujemy:

	20 obserwacji	100 obserwacji
dolna granica	1.793	1.731
górna granica	6.612	3.030

	20 obserwacji	100 obserwacji
prawdopodobieństwo	0.598	0.564

Widzimy, że zgodnie z przewidywaniami, dla większej próby uzyskiwane przedziały są dokładniejsze. Poza tym uzyskane wnioski są podobne do tych, uzyskanych dla próby 50 elementowej.