Elementy statystyki - DEST LIO

Zajęcia 4

Estymacja przedziałowa

Zadanie 1. Dane w pliku `Awarie.txt` (patrz Zadanie 3 Zajęcia 2) dotyczą czasu bezawaryjnej pracy (w h.) pięćdziesięciu urządzeń po wykonaniu kapitalnego remontu. Za rozkład teoretyczny czasu bezawaryjnej pracy przyjęto rozkład wykładniczy $Ex(\lambda)$ o gęstości

$$f_\lambda(x) = \lambda \exp(-\lambda x)\, I_{(0,\infty)}(x),$$

gdzie $\lambda > 0$ jest nieznanym parametrem.

(a) Wyznacz $100(1-\alpha)\%$ przedział ufności dla parametru λ. **Wskazówka:** Niech $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ będzie próba prostą z populacji o rozkładzie wykładniczym $Ex(\lambda)$. Wówczas

$$2\lambda n\bar{X} \sim \chi^2(2n),$$

gdzie $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

(b) Oblicz wartość estymatora największej wiarogodności parametru λ ($ENW(\lambda) = 1/\bar{X}$) oraz granice 95% przedziału ufności dla tego parametru (wyznaczonego w części (a)) dla danych z pliku `Awarie.txt`. **Wskazówka:** `eexp()` z pakietu `EnvStats`

```
0.0009079683, (0.0006739116, 0.0011763746)
```

Zadanie 2. Zmienna w pliku `Pomiary.RData` zawiera wyniki 20 pomiarów wielkości błędu (w mm) badanego urządzenia pomiarowego. Zaimportuj dane z pliku `Pomiary.RData`. **Wskazówka:** `load()`

(a) Wykonaj wykres kwantylowy pokazujący stopień zgodności rozkładu empirycznego z rozkładem normalnym. **Wskazówka:** `qqnorm()`, `qqline()` **Odpowiedź:** patrz Rysunek 1

(b) Oszacuj parametry tego rozkładu metodą punktową i przedziałową. **Wskazówka:** `enorm()` z pakietu `EnvStats`
Uwaga: Przedział ufności dla wariancji w modelu normalnym ma postać:

$$\left(\frac{(n-1)s^2}{\chi^2(1-\alpha/2, n-1)}, \frac{(n-1)s^2}{\chi^2(\alpha/2, n-1)}\right),$$

gdzie $\chi^2(\alpha, n)$ jest kwantylem rzędu α z rozkładu $\chi^2(n)$.

```
# dla wartości oczekiwanej
0.124, (-1.284874, 1.532874)
# dla wariancji
9.062036, (5.240989, 19.33176)
```

Zadanie 3. Niech $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie Rayleigha $R(\lambda)$, $\lambda > 0$ jest parametrem.

(a) Napisz funkcję `lambda.cint()` realizującą przybliżony przedział ufności dla parametru λ postaci:

$$\left(\frac{1}{n}\sum_{i=1}^n X_i^2\left(1 - \frac{z(1-\alpha/2)}{\sqrt{n}}\right), \frac{1}{n}\sum_{i=1}^n X_i^2\left(1 + \frac{z(1-\alpha/2)}{\sqrt{n}}\right)\right),$$

gdzie $z(\alpha)$ jest kwantylem rzędu α z rozkładu normalnego $N(0,1)$ (kwantyle te w programie R wyznacza funkcja `qnorm()`). Funkcja powinna posiadać dwa argumenty: `x` - wektor zawierający dane, `conf.level` - poziom ufności. Funkcja zwraca obiekt (listę) klasy `confint` (klasę obiektu w programie R nadaje się funkcją `class()`) o następujących elementach: `title` - nazwa estymowanego parametru, `est` - wartość estymatora największej wiarogodności parametru, `l` - lewy koniec przedziału ufności, `r` - prawy koniec przedziału ufności, `conf.level` - poziom ufności.

(b) Korzystając z funkcji `lambda.cint()`, oblicz wartość estymatora największej wiarogodności parametru λ oraz granice 95% przedziału ufności dla tego parametru dla danych dotyczących średniej szybkości wiatru rozważanych w Zadaniu 1 na Zajęciach 3. **Wskazówka:** Przed wywoływaniem funkcji `lambda.cint()` wczytaj najpierw poniższe przeciążone funkcje `print` i `summary`:

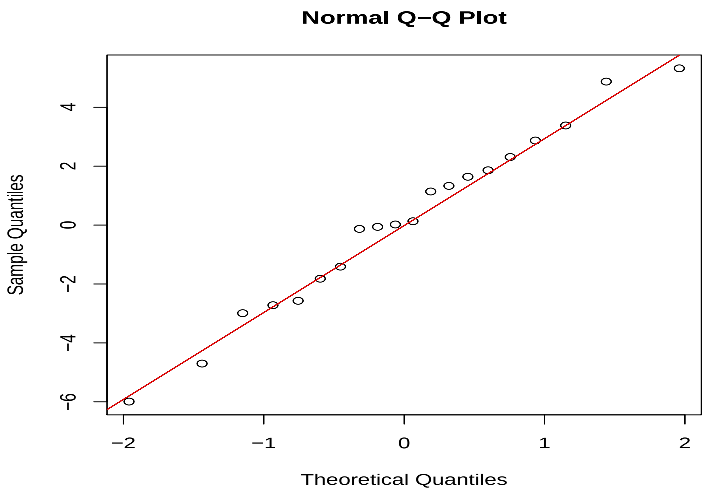
```
print.confint <- function(x){
  cat(x$conf.level*100, "percent confidence interval:", "\n")
  cat(x$l, " ", x$r, "\n")
}
summary.confint <- function(x){
  cat("\n", "Confidence interval of", x$title, "\n", "\n")
  cat(x$conf.level*100, "percent confidence interval:", "\n")
  cat(x$l, " ", x$r, "\n")
  cat("sample estimate", "\n")
  cat(x$est, "\n")
}
```

```
95 percent confidence interval:
21.53562   49.30438
# lub
 Confidence interval of lambda

95 percent confidence interval:
21.53562   49.30438
sample estimate
35.42
```

Zadanie 4. Mamy możliwie dokładnie oszacować zużycie energii elektrycznej w badanym zakładzie w przyszłym miesiącu. Analizowane zużycie energii (w tys. kWh) jest bardzo zmienne i zależy przede wszystkim od wielkości produkcji (w tys. sztuk). Wyniki z poprzednich 10 miesięcy zostały zawarte w pliku `Energia.txt`. Przyjmując model regresji liniowej oraz zakładając normalność rozkładu błędów, wyznacz 95% przedział ufności dla prognozowanego zużycia energii elektrycznej. Planowana w przyszłym miesiącu wielkość produkcji, to 8 tys. sztuk. **Wskazówka:** `lm()`, `predict()`

```
       fit      lwr      upr
1 69.22047 63.51138 74.92957
```



Rysunek 1. Wykres do Zadania 2 (a).