3. PRZEDZIAŁY UFNOŚCI

Definicja. *Przedziałem ufności* dla parametru $\theta \in \Theta \subset \mathbb{R}$ na poziomie ufności $1-\alpha$, gdzie $\alpha \in (0,1)$, nazywamy przedział (θ_1, θ_2), gdzie $\theta_1 = \theta_1(X_1, X_2, \ldots, X_n)$ i $\theta_2 = \theta_2(X_1, X_2, \ldots, X_n)$ to mierzalne funkcje próby takie, że $\theta_1 \leq \theta_2$ i

$$P\left(\theta \in (\theta_1(X_1, X_2, \ldots, X_n), \theta_2(X_1, X_2, \ldots, X_n)\right) = 1-\alpha \ \text{ dla każdego } \ \theta \in \Theta. \tag{1}$$

Końce przedziału ufności $\theta_1(X_1, X_2, \ldots, X_n)$ i $\theta_2(X_1, X_2, \ldots, X_n)$ to zmienne losowe. Dla niektórych realizacji próby losowej otrzymamy przedział $(\theta_1(x_1, x_2, \ldots, x_n), \theta_2(x_1, x_2, \ldots, x_n))$ taki, że

$$\theta \in (\theta_1(x_1, x_2, \ldots, x_n), \theta_2(x_1, x_2, \ldots, x_n)),$$

dla innych realizacji - taki, że

$$\theta \notin (\theta_1(x_1, x_2, \ldots, x_n), \theta_2(x_1, x_2, \ldots, x_n)).$$

Warunek (1) gwarantuje nam, że dla dużej liczby realizacji procent otrzymanych przedziałów takich, że zachodzi $\theta \in (\theta_1(x_1, x_2, \ldots, x_n), \theta_2(x_1, x_2, \ldots, x_n))$ będzie w przybliżeniu równy $(1-\alpha)100\%$.

Podstawowe statystyki próbkowe i funkcje w R wyliczające te statystyki	
średnia z próby: $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$,	wariancja z próby: $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2$
> mean(dane)	> var(dane)

Oznaczenia podstawowych kwantyli i funkcje w R wyliczające te kwantyle		
dla rozkładu normalnego	dla rozkładu t-Studenta	dla rozkładu chi-kwadrat
z_α	$t_{\alpha,n}$	$\chi^2_{\alpha,n}$
> qnorm(α)	> qt(α, n)	> qchisq(α, n)

Przedziały ufności na poziomie ufności $1-\alpha$	
dla wartości średniej μ	**dla wariancji σ^2 (odchylenia standardowego σ)**
Model I. Cecha $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - znane	
$\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	
Model II. Cecha $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - nieznane	
$\overline{x} - t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}$ `> t.test(x,conf.level)$conf.int` gdzie: `x` określa wektor z danymi, `conf.level` określa poziom ufności	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}$
Model III. Cecha X ma rozkład dowolny o skończonej wariancji, próba jest duża ($n \geq 100$)	
$\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$	
Model IV. Cecha X ma rozkład dwupunktowy $P(X=1) = p, P(X=0) = q = 1-p$, p - nieznane	

$\hat{p} = \frac{k}{n} = \frac{\text{ilość sukcesów}}{\text{ilość prób}}$, $\hat{q} = 1-\hat{p}$
`> binom.test(x,n,conf.level)$conf.int`
gdzie: `x` i `n` określają liczbę sukcesów i liczbę prób,
`conf.level` określa poziom ufności.
Jeśli n jest duże i $n\hat{p} > 5$ i $n\hat{q} > 5$, to można wyznaczyć przybliżony przedział asymptotyczny używając:
`> prop.test(x,n,conf.level)$conf.int`	

Wyznaczanie niezbędnej ilości pomiarów do próby do oszcowania wartości średniej μ z maksymalnym błędem d na poziomie ufności (1-α)
Model I. Cecha $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - znane $n \geq \left(z_{1-\alpha/2}\frac{\sigma}{d}\right)^2$
Model II. Cecha $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - nieznane $n \geq \left(t_{1-\alpha/2,n_0-1}\frac{s_0}{d}\right)^2$ gdzie n_0 i s_0 to liczność i odchylenie standardowe pobranej próby wstępnej
Model III. Cecha X ma rozkład dowolny o skończonej wariancji, próba jest duża ($n \geq 100$) $n \geq \left(z_{1-\alpha/2}\frac{s_0}{d}\right)^2$ gdzie s_0 jest odchyleniem standardowym pobranej próby wstępnej
Model IV. Cecha X ma rozkład dwupunktowy $P(X=1) = p, P(X=0) = q = 1-p$, p - nieznane $n \geq z^2_{1-\alpha/2}\frac{p_0 q_0}{d^2}$ jeżeli znany jest szacunkowy procent p_0 (wtedy $q_0 = 1-p_0$) $n \geq z^2_{1-\alpha/2}\frac{1}{4d^2}$ jeżeli nie jest znany szacunkowy procent p_0