Zadanie 6.5

Zadanie 6.5. Jak dużą próbę należy pobrać, aby z maksymalnym błędem 2% oszacować na poziomie ufności 0.99 odsetek kierowców niezapinających pasów bezpieczeństwa? Uwzględnić rezultaty wstępnych badań, z których wynika, że interesująca nas wielkość jest rzędu 16%. Porównać otrzymaną liczność próbki z licznością, jaka byłaby wymagana, gdyby pominąć rezultaty badań wstępnych.

Próbka $X_1, ..., X_n$ z rozkładu $Ber(\theta)$. Przedział ufności

$$CI_{1-\alpha}(\theta) = \left(\hat{\theta} - z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}, \hat{\theta} + z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\right)$$

to jest p. ufności Walda, dla p-stwa sukcesu

zapina / nie zapina

Błąd

tą funkcję tutaj możemy z gdy

$$z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \leqslant z_{1-\alpha/2}\sqrt{\frac{0.5(1-0.5)}{n}} = \boxed{z_{1-\alpha/2}\sqrt{\frac{1}{4n}}}$$

w najgorszym wypadku ten błąd nie zależy od θ

Stąd gdy mamy badanie wstępne i $\hat{\theta}_0 = 0.16$, szukamy n takiego, że

tu n uwzględni badanie wstępne →

$$z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}_0(1-\hat{\theta}_0)}{n}} \leqslant 0.02$$

Gdy nie mamy, szukamy n takiego, że

a tutaj nie →

$$z_{1-\alpha/2}\sqrt{\frac{1}{4n}} \leqslant 0.02$$

Zadanie 6.10

Mamy dwie próby $X_1, ..., X_{n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ i $Y_1, ..., Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Szukamy $CI_{1-\alpha}(\sigma_1^2/\sigma_2^2)$. Dla przypomnienia:

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Kartka 2.

oraz, gdy $X \sim \chi^2_n$ i $Y \sim \chi^2_m$, to

$$\boxed{\frac{X/n}{Y/m} \sim F^{[n,m]}}$$

Kartka 2
F- snedecor

Stąd

$$Q(\mathbb{X}, \mathbb{Y}, \sigma_1^2/\sigma_2^2) = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F^{[n_1-1, n_2-1]}$$

A więc

$$P(F_{\alpha/2}^{[n_1-1,n_2-1]} \leqslant \frac{S_X^2}{S_Y^2}\frac{\sigma_2^2}{\sigma_1^2} \leqslant F_{1-\alpha/2}^{[n_1-1,n_2-1]}) = 1 - \alpha$$

Zadanie 6.10

relacja między kwantylami

W końcu

$$CI_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left(\frac{S_X^2}{S_Y^2} \frac{1}{F_{1-\alpha/2}^{[n_1-1,n_2-1]}}, \frac{S_X^2}{S_Y^2} \frac{1}{F_{\alpha/2}^{[n_1-1,n_2-1]}} \right)$$

Lub alternatywnie:

$$CI_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left(\frac{S_X^2}{S_Y^2} F_{\alpha/2}^{[n_2-1,n_1-1]}, \frac{S_X^2}{S_Y^2} F_{1-\alpha/2}^{[n_2-1,n_1-1]} \right)$$

Przykład: $n_1 = 10$, $n_2 = 7$, $S_X^2 = 2.1$, $S_Y^2 = 3.5$, $F_{0.025}^{[6,9]} = 0.18$, $F_{0.975}^{[6,9]} = 4.32$

$$CI_{1-\alpha}(\sigma_1^2/\sigma_2^2) = (0.108, 2.586)$$

Zadanie 6.18

Mamy $X_1, ..., X_n \sim Ber(\theta)$. Szukamy asymptotycznego $CI_{1-\alpha}(\theta)$. Przypomnijmy, że

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i}{n}$$

Przedział ufności Wilsona (ang. *score confidence interval*)

Ponieważ $\frac{\widehat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \underset{n\to\infty}{\sim} \mathrm{N}(0,1)$, więc dla dostatecznie licznej próbki mamy

$$\mathbb{P}\Big(|\widehat{\theta}-\theta|\sqrt{\frac{n}{\theta(1-\theta)}} < z_{1-\alpha/2}\Big) \simeq 1-\alpha,$$

Z CLT mamy, że dla dużego n:

$$Q(\mathbb{X}, \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)/n}} \approx \mathcal{N}(0, 1)$$

A stąd

$$P(-z_{1-\alpha/2} \leqslant \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n} \leqslant z_{1-\alpha/2}) \approx 1 - \alpha.$$

lub równoznacznie

$$P(\frac{|\hat{\theta} - \theta|}{\sqrt{\theta(1-\theta)}}\sqrt{n} \leqslant z_{1-\alpha/2}) \approx 1 - \alpha.$$

Zadanie 6.18

A więc musimy znaleźć θ, które spełnia nierówność:

$$(\hat{\theta} - \theta)^2 n \leqslant z^2_{1-\alpha/2}\theta(1-\theta)$$

Porządkując, otrzymamy

$$\theta^2(n + z^2_{1-\alpha/2}) - \theta(2\hat{\theta}n + z^2_{1-\alpha/2}) + \hat{\theta}^2 n \leqslant 0$$

Dalej rozwiążmy ... = 0

$$\Delta = b^2 - 4ac = z^4_{1-\alpha/2} + 4nz^2_{1-\alpha/2}\hat{\theta}(1-\hat{\theta})$$

Stąd

$$\theta_{1/2} = \frac{-b \pm \sqrt{\Delta}}{2a} = \frac{2\hat{\theta}n + z^2_{1-\alpha/2} \pm \sqrt{z^4_{1-\alpha/2} + 4nz^2_{1-\alpha/2}\hat{\theta}(1-\hat{\theta})}}{2(n + z^2_{1-\alpha/2})}$$

A stąd

$$CI_{1-\alpha}(\theta) = (\theta_1, \theta_2)$$

jest bardziej skomplikowany niż Wald, ale trochę dokładniejszy