Mathematical Statistics

Class 6. Confidence Intervals: Estimation of population mean given variance (problems), estimation of population proportion.

MDI, September 2022.

Estimating population proportion. Large sample.

Let's assume we have random sample: $X_1, \ldots, X_n$, with k positive answers, where each X_i is Bernoulli random variable with probability of success p_1, $n > 30$. We are interested in estimation of the population parameter p — population proportion.

We introduce a point estimator $\hat{p} = \frac{k}{n}$, which we call a sample proportion.

If $n > 30$ then, as a consequence of the *Central Limit Theorem* we have:

$$\boxed{\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)} \tag{1}$$

Then, classic procedure:

$$\begin{aligned} 1-\alpha = P(L < p < U) = P(-U < -p < -L) = \\ = P\left(\frac{\hat{p}-U}{\operatorname{Var}(\hat{p})} < \frac{\hat{p}-p}{\operatorname{Var}(\hat{p})} < \frac{\hat{p}-L}{\operatorname{Var}(\hat{p})}\right) \end{aligned}$$

If all necessary conditions are fulfilled, and Eq. (1) is true, then the fraction $\frac{\hat{p}-p}{\operatorname{Var}(\hat{p})}$ behaves as Standard Normal random variable $Z \sim \mathcal{N}(0,1)$. So we can rewrite the last equation as:

$$1-\alpha = P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right).$$

We find constant $z_{\alpha/2}$ from the statistical table, according to our choice of confidence level. After that is done, we can write down bounds for required confidence interval:

$$\begin{aligned} L &= \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ U &= \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \end{aligned}$$

where we change p to its point estimate $\hat{p}$, because we do not know the true parameter, and sample proportion is the only thing we have in disposal.

The $(1-\alpha)100\%$ Confidence Interval for the difference of population proportions:

$$\boxed{p \in (L, U)} \tag{2}$$

Problems

1. Soon after he took office in 1963, President Johnson was approved by 160 out of a sample of 200 Americans. With growing disillusionment over his Vietnam policy, by 1968 he was approved by only 70 out of a sample of 200 Americans. What is the 95% confidence interval for the percentage of all Americans who approved Johnson in 1968? In 1963?