Engineering Mathematics and Statistics (B39AX)
Fall 2023

Tutorial 5

Problem A. Last year, a telephone company in Scotland undertook a survey of telephone usage. According to the media relations manager, the company randomly selected $n = 15000$ local telephone calls of residential customers in Glasgow. The mean duration of the sampled calls was $\overline{x} = 3.8$ minutes. Use this information to determine a 95% confidence interval for the true mean duration μ of all telephone calls made by residential customers in Glasgow. Assume each call has a standard deviation of $\sigma = 4.0$ minutes.

Let X_i be the random variable representing the duration of call i, $\mu = \mathbb{E}[X_i]$, and $\mathrm{Var}(X_i) = \sigma^2$. Assuming the calls are independent, and noticing that $n = 15000 \gg 30$, we can use the central limit theorem and consider that

$$Z := \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\,,$$

where the sample mean $\overline{X}_n = (1/n)\sum_{i=1}^{n} X_i$ has normal distribution $\mathcal{N}(\mu, \sigma^2/n)$, and $Z \sim \mathcal{N}(0,\,1)$ a standard normal distribution.

A 95% confidence interval for μ has the form $\mathbb{P}(\Theta_n^- \leq \mu \leq \Theta_n^+) \geq 0.95$, where Θ_n^- and Θ_n^+ are random variables that depend on the observations of the data. To obtain such confidence interval for μ, notice

$$\mathbb{P}\big(-c \leq Z \leq c\big) = \mathbb{P}\bigg(-c \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\bigg) = \mathbb{P}\bigg(\underbrace{\overline{X}_n - c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^-} \leq \mu \leq \underbrace{\overline{X}_n + c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^+}\bigg)\,.$$

We then just need to determine c such that $\mathbb{P}(-c \leq Z \leq c) \geq 0.95$ or, by visual inspection of the pdf of Z, $\Phi(c) \geq 0.025 + 0.95 = 0.975$. Using the table for the standard normal distribution, $c = 1.96$ will do. Replacing c, σ, n, and the realization $\overline{x}_n = 3.8$ of $\overline{X}_n$ in Θ_n^- and Θ_n^+, we obtain the 95% confidence interval

$$\bigg[3.8 - 1.96\frac{4.0}{\sqrt{15000}},\, 3.8 + 1.96\frac{4.0}{\sqrt{15000}}\bigg] = [3.74,\;3.86]\,.$$

Problem B. A car manufacturer wants to find the average consumption of a new car, measured in km/L. From previous experience, he knows that the standard deviation is 3.0 km/L. How many trials does the manufacturer need to run to be able to state with 99% confidence that he knows the average to within 0.5 km/L?

Let X_i represent the consumption of the car in km/L during trial i. Let $\mu = \mathbb{E}[X_i]$ and $\mathrm{Var}(X_i) = \sigma^2 = 3^2 = 9$. Assuming that each trial is independent (for example, different courses) and that we will obtain $n \geq 30$, we can use the central limit theorem and consider that

$$Z := \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\,,$$

where $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$, has standard normal distribution $\mathcal{N}(0,1)$.
We want to compute n such that $\mathbb{P}(\Theta_n^- \leq \mu \leq \Theta_n^+) \geq 0.99$ and $\theta_n^+ - \theta_n^- = 0.5 \times 2 = 1$. We have

$$\mathbb{P}(-c \leq Z \leq c) = \mathbb{P}\Big(-c \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\Big) = \mathbb{P}\Big(\underbrace{\overline{X}_n - c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^-} \leq \mu \leq \underbrace{\overline{X} + c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^+}\Big).$$

To determine c, use the tables to obtain $\Phi(c) \geq 0.99 + 0.005 = 0.995$. For example, $c = 2.58$ will do. Setting $\theta_n^+ - \theta_n^- = 1$, we obtain

$$1 = 2\,c\frac{\sigma}{\sqrt{n}} \qquad \Longleftrightarrow \qquad n = 4c^2\sigma^2\,.$$

Replacing c and σ, $n = 239.6$, so 240 trials will be required. Notice that $n \geq 30$, which conforms to our use of the central limit theorem.

Problem C. A manufacturer of car tyres needs to estimate the mean lifetime μ of a new line of steel-belted radials. From past experience, it is known that the population standard deviation of tyre lifetime is $\sigma = 2500$ miles, and that the tyre lifetime is normally distributed. The results from independent tests on a random sample of $n = 16$ tyres are displayed below.

40133	37494	39446	42294
39433	40403	41559	37176
38309	41224	35012	39322
40572	38544	40882	39704

Use the data to obtain a 90% confidence interval for the true mean life μ of this new line of steel-belted radial. Note that $\sum_{i=1}^{16} x_i = 631507$.

Let X_i denote the lifetime of tyre i, $\mu = \mathbb{E}[X_i]$, and $\mathrm{Var}(X_i) = \sigma^2 = 2500^2$. In this example, $n = 16 < 30$ so, if we did not know the distribution of each tyre's lifetime so we could not use the central limit theorem.
We will then use the fact that the sum of normal random variables has normal distribution. Specifically, if $X_i \sim \mathcal{N}(\mu, \sigma^2)$, for $i = 1, \ldots, n$, and the X_i's are independent, then $X_1 + \cdots + X_n \sim \mathcal{N}(n\,\mu, n\,\sigma^2)$. Using the properties of the expected value and variance,

$$\overline{X}_n := \frac{X_1 + \cdots + X_n}{n} \sim \mathcal{N}(\mu,\, \sigma^2/n)\,, \quad \text{and} \quad Z := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,\,1)\,.$$

To obtain a 90% confidence interval for μ, we compute c such that

$$\mathbb{P}(-c \leq Z \leq c) = \mathbb{P}\Big(-c \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\Big) = \mathbb{P}\Big(\underbrace{\overline{X}_n - c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^-} \leq \mu \leq \underbrace{\overline{X} + c\frac{\sigma}{\sqrt{n}}}_{\Theta_n^+}\Big) \geq 0.9\,.$$

Using the symmetry of $\mathcal{N}(0,1)$, $\Phi(c) \geq 0.9 + 0.05 = 0.95$. Using the tables, $c = 1.65$ will do. Replacing the values in the expressions for Θ_n^- and Θ_n^+, we obtain the 90% confidence interval

$$\left[\overline{x}_n - c\frac{\sigma}{\sqrt{n}},\, \overline{x}_n + c\frac{\sigma}{\sqrt{n}}\right] = \left[\frac{631507}{16} - 1.65\,\frac{2500}{4},\, \frac{631507}{16} + 1.65\,\frac{2500}{4}\right] \simeq [38438,\, 40500]\,.$$