CE2100/CZ2100
Probability and Statistics for Computing

Tutorial 2 - Large-Sample Estimation

Problem 1

An increase in the rate of consumer savings is frequently tied to a lack of confidence in the economy. A random sample of $n = 200$ savings accounts in a local community showed a mean increase in savings account values of 7.2% over the past 12 months, with a standard deviation of 5.6%. Estimate the mean percentage increase in savings account values over the past 12 months for depositors in the community. Find the margin of error for your estimate.

Solution:

The point estimation of μ is $\bar{x} = 7.2\%$ and the margin of error can be calculated by

$$1.96SE = 1.96\frac{\sigma}{\sqrt{n}} \approx 0.776$$

Problem 2

An opinion poll indicated that 49% of the 1034 adults surveyed think that country A should pursue a program to send humans to Mars. Estimate the true proportion of people who think that country A should pursue this program. Calculate the margin of error.

Solution:

The point estimation of p is $\hat{p} = \frac{x}{1034} = 0.49$, and the margin of error is computed as

$$1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} \approx 0.0305$$

Problem 3

Each of $n = 30$ students in a chemistry class measured the amount of copper precipitated from a saturated solution of copper sulfate over a 30-minute period. The sample mean and standard deviation of the 30 measurements were equal to 0.145 and 0.0051 mole, respectively. Find a 90% confidence interval for the mean amount of copper precipitated from the solution over a 30-minute period.

Solution:

With $n = 30$, $\bar{x} = 0.145$ and $s = 0.0051$, a 90% confidence interval for μ is approximated by

$$\bar{x} \pm 1.645 \frac{s}{\sqrt{n}} = 0.145 \pm 0.0015$$

which is $0.1435 < \mu < 0.1465$.

Problem 4

A survey is designed to estimate the proportion of sports utility vehicles (SUVs) being driven in the state of California. A random sample of 500 registrations is selected from a Department of Motor Vehicles database, and 68 are classified as SUVs. Use a 95% confidence interval to estimate the proportion of SUVs in California.

Solution:

The point estimation of p is $\hat{p} = \frac{x}{n} = \frac{68}{500} = 0.136$, and the approximate 95% confidence interval for p is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.136 \pm 0.030$$

or $0.106 < p < 0.166$.

Problem 5

It is reported that, in a study of a particular wafer inspection process, 356 dies were examined by an inspection probe and 201 of these passed the probe. Assuming a stable process, calculate a 95% (two-sided) confidence interval for the proportion of all dies that pass the probe.

Solution:

Note that

$$\hat{p} = \frac{201}{356} = 0.5646$$

We then calculate the 95% confidence interval as

$$0.5646 \pm 1.96\sqrt{\frac{0.5646 \times 0.4354}{356}} = (0.513, 0.616)$$

Problem 6

For a study, we conduct on nutrition and access to fresh produce in Beaufort County, North Carolina. We want to know how much an adult spends on locally-produced fruit and vegetables in June. We randomly select 100 individuals from the county property records and send a survey to those residents about their eating, shopping and gardening practices. With our sample, we find that the average amount an adult spends on locally-grown fruits and vegetables in June is \$40.00. We know from previous studies that the standard deviation of money spent on local produce is \$10. Construct a 95% confidence interval for the mean (per capita) amount spent on fresh, local produce.

Solution:

To construct our confidence interval, we know that the sample mean is \$40.00 and the population standard deviation is \$10. Our sample size is 100. The z^* value we will use is 1.96. Therefore, the confidence interval can be calculated as:

$$(40.00 - 1.96 \cdot \frac{10}{\sqrt{100}}, 40.00 + 1.96 \cdot \frac{10}{\sqrt{100}}) = (38.04, 41.96)$$

Additional Questions (Do not discuss in tutorial)

Problem 7

It is reported that for a sample of 50 kitchens with gas cooking appliances monitored during a one-week period, the sample mean of CO_2 level (ppm) was 654.16, and the sample standard deviation was 164.43.

(a) Calculate the 95% (two-sided) confidence interval for the true average CO_2 level in the population of all homes from which the sample was selected.

(b) Suppose the investigators had made a rough guess of 175 for the value of standard deviation s before collecting data. What sample size would be necessary to obtain an interval width of 50 ppm for a confidence level of 95%?

Solution:

(a)

$$\bar{x} \pm z_{0.25} \frac{s}{\sqrt{n}} = 654.16 \pm 1.96 \frac{164.43}{\sqrt{50}} = (608.58, 699.74)$$

Thus, the confidence interval is $(608.58, 699.74)$.

(b)

$$w = 50 = \frac{2(1.96)(175)}{\sqrt{n}} \to \sqrt{n} = 13.72$$

Thus,

$$n = 188.24$$

which rounds up to 189.

Problem 8

Let $X_1, \dots, X_n$ be a random sample from population with mean μ and variance σ^2. Show that the sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an unbiased estimator for σ^2.

Solution:

To show that S^2 is unbiased, we need to show that

$$\mathbb{E}(S^2) = \mathbb{E}\left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2\right) = \sigma^2$$

Note that

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) &= \mathbb{E}(X^2 - 2\mu X + \mu^2)\\ &= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mathbb{E}(\mu^2)\\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2\\ &= \mathbb{E}(X^2) - \mu^2\end{aligned}$$

Thus, $\mathbb{E}(X^2) = \mathrm{Var}(X) + \mathbb{E}(X)^2 = \sigma^2 + \mu^2$. Using this result, we can also derive

$$\mathbb{E}(\bar{X}^2) = \mathrm{Var}(\bar{X}) + \mathbb{E}(\bar{X})^2 = \sigma^2/n + \mu^2,$$

since $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$ and $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu$.

Therefore,

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) &= \mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2\right)\\ &= \mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2\bar{X}\left(n\frac{1}{n}\sum_{i=1}^n X_i\right) + n\bar{X}^2\right)\\ &= \mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right)\\ &= \sum_{i=1}^n \mathbb{E}(X_i^2) - \mathbb{E}\left(n\bar{X}^2\right)\\ &= \sum_{i=1}^n \mathbb{E}(X_i^2) - n\mathbb{E}\left(\bar{X}^2\right)\\ &= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2\end{aligned}$$

Thus, $\mathbb{E}(S^2) = \sigma^2$.