HW4

Smetanin Aleksandr

27 марта 2020 г.

$X_1, \dots X_n$ - independent equally distributed random variables with expected value μ and standard deviation σ. In case we don't know the distribution of a random variable, we use central limit theorem (CLT) to build confidence interval:

$$\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1), n \to \infty$$

$$P(-t_{\frac{1+\gamma}{2}} < \sqrt{n}\frac{\bar{X}-\mu}{\sigma} < t_{\frac{1+\gamma}{2}}) = \gamma$$

$$P(\bar{X} - t_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = \gamma$$

Margin of error with confidence γ:

$$MOE_\gamma = t_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

1. The 92 million Americans of age 50 and over control 50 percent of all discretionary income (AARP Bulletin, March 2008). AARP estimated that the average annual expenditure on restaurants and carryout food was \$1873 for individuals in this age group. Suppose this estimate is based on a sample of 80 persons and that the sample standard deviation is \$550.
a) At 95% confidence, what is the margin of error?

$$MOE_{0.95} = \tau_{0.975}\frac{\bar{\sigma}}{\sqrt{n}} = 1.96 * \frac{550}{\sqrt{80}} = 121\$$$

b) What is the 95% confidence interval for the population mean amount spent on restaurants and carryout food?

$$(\bar{X} - MOE_{0.95}, \bar{X} + MOE_{0.95})$$
$$(1752, 1994)$$

c) What is your estimate of the total amount spent by Americans of age 50 and over on restaurants and carryout food?

$$(1752 \cdot 92 \cdot 10^6, 1994 \cdot 92 \cdot 10^6) = (161 \cdot 10^9, 183 \cdot 10^9)$$

d) If the amount spent on restaurants and carryout food is skewed to the right, would you expect the median amount spent to be greater or less than \$1873?
Greater.

2. Mileage tests are conducted for a particular model of automobile. If a 98% confidence interval with a margin of error of 1 mile per gallon is desired, how many automobiles should be used in the test? Assume that preliminary mileage tests indicate the standard deviation is 2.6 miles per gallon.

$$P(-\tau_{0.99} < \sqrt{n}\frac{\bar{X}-\mu}{\bar{\sigma}} < \tau_{0.99}) \xrightarrow[n\to\infty]{} 0.98$$

$$P(\bar{X} - \tau_{0.99}\frac{\bar{\sigma}}{\sqrt{n}} < \mu < \bar{X} - \tau_{0.99}\frac{\bar{\sigma}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0.98$$

$$MOE_{0.98} = \tau_{0.99}\frac{\bar{\sigma}}{\sqrt{n}}$$

$$n = \left(\frac{\tau_{0.99}\bar{\sigma}}{MOE_{0.98}}\right)^2 = \left(\frac{2.326 \cdot 2.6}{1}\right)^2 = 37$$

3. In developing patient appointment schedules, a medical center wants to estimate the mean time that a staff member spends with each patient. Use a planning value for the population standard deviation of eight minutes.
a) How large a sample should be taken if the desired margin of error is two minutes at a 95% level of confidence?

$$MOE_{0.95} = \tau_{0.975}\frac{\sqrt{\bar{S}_n}}{\sqrt{n}}$$

$$n = \left(\frac{\tau_{0.975} \cdot \bar{\sigma}}{MOE_{0.95}}\right)^2 = \left(\frac{1.96 \cdot 8}{2}\right)^2 = 64$$

b) How large a sample should be taken for a 99% level of confidence?

$$n = \left(\frac{\tau_{0.995} \cdot \bar{\sigma}}{MOE_{0.99}}\right)^2 = \left(\frac{2.576 \cdot 8}{2}\right)^2 = 106$$

4. The 2003 Statistical Abstract of the United States reported the percentage of people 18 years of age and older who smoke. Suppose that a study designed to collect new data on smokers and nonsmokers uses a preliminary estimate of the proportion who smoke of .30.
a) How large a sample should be taken to estimate the proportion of smokers in the population with a margin of error of .02? Use 95% confidence.

$$P(-\tau_{0.975} < \sqrt{n}\frac{\bar{X} - p}{\sqrt{\bar{S}_n}} < \tau_{0.975}) \xrightarrow[n\to\infty]{} 0.95$$

$$MOE_{0.95} = \tau\sqrt{\frac{\bar{S}_n}{n}}, n \to \infty$$

$$n = \left(\frac{\tau_{0.975}}{MOE_{0.95}}\right)^2 \bar{S}_n = \left(\frac{1.96}{0.02}\right)^2 \cdot 0.3 \cdot 0.7 = 2017$$

b) Assume that the study uses your sample size recommendation in part (a) and finds 520 smokers. What is the point estimate of the proportion of smokers in the population?

$$\hat{p} = \bar{X} = \frac{520}{2017} = 0.258$$

c) What is the 95% confidence interval for the proportion of smokers in the population?

$$P(\bar{X} - \tau_{0.975}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \bar{X} - \tau_{0.975}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0.95$$

$$(\bar{X} - \tau_{0.975}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \bar{X} + \tau_{0.975}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}})$$

$$(0.258 - 1.96 \cdot \sqrt{\frac{0.258 \cdot 0.742}{2017}}, 0.258 + 1.96 \cdot \sqrt{\frac{0.258 \cdot 0.742}{2017}})$$

$$(0.239, 0.277)$$

5. Let $X_n \sim Bin(n, p)$, where $p \in (0, 1)$ is unknown. Since $\sqrt{n}(\frac{X_n}{n} - p) \Rightarrow N(0, p(1-p))$, the variance of the limiting distribution depends only on p. Derive the confidence interval for p using the variance-stabilizing transformation.

$$\sigma^2(\mu)(g'(\mu))^2 = 1 \Rightarrow g'(\mu) = \frac{1}{\sigma(\mu)}$$

$$g'(p) = \frac{1}{\sqrt{p(1-p)}}$$

$$g(p) = -\arcsin(1-2p)$$

$$\sqrt{n}\left(g\left(\frac{X_n}{n}\right) - g(p)\right) \Rightarrow N(0,1)$$

$$P\left(-\tau_{\frac{1+\gamma}{2}} < \sqrt{n}\left(\arcsin(1-2p) - \arcsin\left(1-2\frac{X_n}{n}\right)\right) < \tau_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(\arcsin\left(1-2\frac{X_n}{n}\right) - \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}} < \arcsin(1-2p) < \arcsin\left(1-2\frac{X_n}{n}\right) + \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right) = \gamma$$

$$P\left(1-\sin\left(\arcsin\left(1-2\frac{X_n}{n}\right) + \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right) < 2p < 1-\sin\left(\arcsin\left(1-2\frac{X_n}{n}\right) - \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right)\right) = \gamma$$

Interval for p with confidence level γ

$$\left(\frac{1-\sin\left(\arcsin\left(1-2\frac{X_n}{n}\right) + \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right)}{2}, \frac{1-\sin\left(\arcsin\left(1-2\frac{X_n}{n}\right) - \frac{\tau_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right)}{2}\right).$$