Discussion #6

Review: Confidence Interval

1. Suppose $X \sim N(\mu, \sigma^2)$, and $X_1, X_2, ..., X_n$ is a random sample from this distribution.

   (a) If σ is known, the $(1-\alpha)$ C.I. for μ is

   $$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

   where $Z_{\alpha/2}$ is such that
   $P(Z \ge Z_{\alpha/2}) = \alpha/2$

   (b) If σ is unknown, then the $(1-\alpha)$ C.I. for μ is

   $$\bar{x} - T_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + T_{n-1,\alpha/2}\frac{s}{\sqrt{n}},$$

   where $T_{n-1,\alpha/2}$ is such that
   $P(T_{n-1} \ge T_{n-1,\alpha/2}) = \alpha/2$

   (c) The $(1-\alpha)$ C.I. for σ^2 is

   $$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}},$$

   For testing the variance, use

   $$V^2 = \frac{(n-1)S^2}{\sigma^2}$$

   where σ^2 is the hypothesized value in H_o. V^2 is exactly χ^2_{n-1} distribution.

2. In general, if E(X)= $\mu, Var(X) = \sigma^2$, and sample size n is large, then the $(1-\alpha)$ C.I. for μ is

   $$\bar{x} - Z_{\alpha/2}\ \sigma_{\bar{X}} \le \mu \le \bar{x} + Z_{\alpha/2}\ \sigma_{\bar{X}}$$

   where $\sigma_{\bar{X}} = \begin{cases} \sigma/\sqrt{n} & \text{if } \sigma \text{ is known} \\ s/\sqrt{n} & \text{if } \sigma \text{ is unknown} \end{cases}$

3. **Relation between C.I. and two-sided hypothesis testing** (not for binomial distribution):
   If the $(1-\alpha)$ C.I. for μ contains the hypothesized value μ_0, then we do not reject the null hypothesis H_0 at level α. Otherwise we reject H_0 at level α.

4. If X is distributed as B(n, p), and n is reasonably large, then the $(1-\alpha)$ C.I. for p is

   $$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

   Note that in hypothesis testing, we use the hypothesized value p_0 to calculate p-value, but we use $\hat{p}$ in computing C.I. for p.

Practice Problem

Review Problem from Chapter 6: Inference on Variance

1. It is thought that the variability in measuring the weights of watermelon seeds is at most 0.0025 gm^2. You obtain the following random sample of 8 watermelon seeds: 0.82 0.71 0.77 0.67 0.70 0.85 0.73 0.77. Test the claim and evaluate the assumptions required for this test.

   Problems for Chapter 7: Confidence Intervals

2. The time to blossom of 21 plants has $\bar{X} = 39$ days and $s = 5.1$ days. Assume that the time of blossom of a plant is normally distributed.

   (a) Give a 90% C.I. of the mean time to blossom.

   (b) Describe the effects on the C.I. if σ is known and $\sigma = 5.1$.

   (c) Compute a 90% C.I. for the population variance of the time to blossom.

3. A forester measures 100 needles of a pine tree and finds $\bar{X} = 3.1$ cm and $s = 0.7$ cm.

   (a) What is a point estimate for μ i.e., the population mean of needle length of the tree ?

   (b) Construct 95% C.I. for μ. Does the interval cover the true mean ?

   (c) Test the claim that $\mu = 3$ against the 2 sided alternative using result in (b).

4. A machine in a food processing factory must be repaired if it produces more than 10% defectives among the large lot of items it produces in a day. A random sample of 100 items from the day's production contains 15 defectives.

   (a) Using $\alpha = 0.01$, would you conclude that the machine needs repair ?

   (b) Find 99% C.I. for the % defectives in the population.

   (c) State (if any) the relationship between results in (a) and (b).

5. Suppose we are sampling from $N(\mu, 64)$ distribution. How large must n be so that a 95% C.I. for μ has length equal to 0.1?

Solution

1. Assume normal data, $n = 8$, $\bar{x} = 0.7525$, $S^2 = 0.00379$
$H_0: \ \sigma^2 = 0.0025 \ \ vs \ \ H_A: \ \sigma^2 > 0.0025$

$$\text{p-value} = P(S^2 \geq 0.00379|H_0)$$
$$= P(\frac{(n-1)S^2}{\sigma^2} \geq \frac{7 \times 0.00379}{0.0025})$$
$$= P(V^2 \geq 10.62)$$

From the chi-squared tables, we determine that $0.1 < p - value < 0.25$. Hence there is no meaningful evidence against H_0.

2. Normal data, $\bar{x} = 39, s = 5.1, n = 21$.

   (a) $\alpha/2 = 0.05$,
$$\bar{x} - T_{20,0.05}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + T_{20,0.05}\frac{s}{\sqrt{n}}$$
$$39 - 1.725\frac{5.1}{\sqrt{21}} \leq \mu \leq 39 + 1.725\frac{5.1}{\sqrt{21}}$$
$$37.08 \leq \mu \leq 40.92$$

   (b) $\sigma = 5.1$,
$$\bar{x} - Z_{0.05}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{0.05}\frac{\sigma}{\sqrt{n}}$$
$$39 - 1.645\frac{5.1}{\sqrt{21}} \leq \mu \leq 39 + 1.645\frac{5.1}{\sqrt{21}}$$
$$37.17 \leq \mu \leq 40.83$$
The C.I. in part(b) is slightly narrower than that in part(a).

   (c) $\alpha/2 = 0.05, 1 - \alpha/2 = 0.95$
$$\frac{(n-1)s^2}{\chi^2_{20,0.05}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{20,0.95}}$$
$$\frac{20(5.1)^2}{31.41} \leq \sigma^2 \leq \frac{20(5.1)^2}{10.85}$$
$$16.56 \leq \sigma^2 \leq 47.94 \ \ or \ \ 4.07 \leq \sigma \leq 6.92$$

3. (a) A point estimate for μ is 3.1.

   (b) $\alpha/2 = 0.025$, $n = 100 > 30$.
$$\bar{x} - Z_{0.025}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{0.025}\frac{s}{\sqrt{n}}$$
$$3.1 - 1.96\frac{0.7}{\sqrt{100}} \leq \mu \leq 3.1 + 1.96\frac{0.7}{\sqrt{100}}$$
$$2.9628 \leq \mu \leq 3.2372$$

   (c) $H_0: \ \mu = 3 \ \ vs. \ \ H_A: \ \mu \neq 3$
Since the $1 - 0.05 = 95\%$ confidence interval for μ contains 3, we don't reject the null hypothesis H_0 at 0.05 level.

4. Let X = number of defectives, then $X \sim Bi(100, p)$

   (a) $H_0: \ p = 0.1 \ \ vs \ \ H_A: \ p > 0.1$
$$\begin{aligned} p - value &= P(X \geq 15|H_0) \\ &= P(\frac{X_{NA} - \mu_X}{\sigma_X} \geq \frac{15 - \mu_X}{\sigma_X}) \\ &= P(Z \geq \frac{15 - 100 \times 0.1}{\sqrt{100 \times 0.1 \times 0.9}}) \\ &= P(Z \geq 1.67) = 0.0475 \end{aligned}$$

   (b) $\alpha/2 = 0.01/2 = 0.005$, $\hat{p} = X/n = 15/100 = 0.15$
$$\hat{p} - Z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$0.15 - 2.576\sqrt{\frac{0.15 \times (1 - 0.15)}{100}} \leq p$$
and $p \leq 0.15 + 2.576\sqrt{\frac{0.15 \times (1 - 0.15)}{100}}$
$$0.058 \leq p \leq 0.242$$

   (c) The direct relationship between confidence intervals and significant testing does not hold for confidences intervals obtained from the normal approximation to the binomial.

5. normal data, $\sigma = 8$, $\alpha = 0.025$
$$\bar{x} - Z_{0.025}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{0.025}\frac{\sigma}{\sqrt{n}}$$
$$\text{C.I. length} = (\bar{x} + Z_{0.025}\frac{\sigma}{\sqrt{n}}) - (\bar{x} - Z_{0.025}\frac{\sigma}{\sqrt{n}})$$
$$= 2Z_{0.025}\frac{\sigma}{\sqrt{n}} = 2 \times 1.96 \times \frac{8}{\sqrt{n}} = 0.1$$
Thus $\sqrt{n} = 313.6$ and $n = 98345$.