MAT2377

Catalin Rada

Version June 2, 2009

8-3 Estimating the mean of a normal population (σ unknown)

In the special case of a normal population, it is possible to construct a C.I. for the mean even when is σ unknown.

Definition: Consider the random variable T with probability density function

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\Gamma(\nu/2)}[(t^2/\nu)+1]^{-(\nu+1)/2}$$

for $-\infty < t < \infty$, where $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$ for $r > 0$ is the gamma function. Properties of gamma function:

— when r is integer, $\Gamma(r) = (r-1)!$, so $\Gamma(1) = 0! = 1$; $\Gamma(1/2) = \pi^{1/2}$.

We say that T follows a t distribution with ν degrees of freedom. The mean and variance of T are zero and $k/(k-2)$, for $k > 2$. More Properties: Consider a t distribution with ν degrees of freedom.

— as for the normal we define the percentile $t_{\alpha,\nu}$; and we find them in Table V;

— the density $f(t)$ is symmetric about $t = 0$, hence $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$;

— if $\nu \to \infty$, then the t distribution is a standard normal, hence $z_\alpha = t_{\alpha,\infty}$. So we can use table V to find percentiles for the $N(0, 1)$ distribution. For example, $z_{0.025} = t_{0.025,\infty} = 1.96$.

Theorem: Let $X_1, ..., X_n$ is a random sample from a normal population with mean μ and variance σ^2. The random variable $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ has a t distribution with $\nu = n-1$ degrees of freedom (S^2 is the sample variance).

Therefore, if the population is normal, then a $100(1-\alpha)\%$ confidence interval for μ is: $\overline{x} \pm t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$.

EXC: A SuperCola beverage machine is made up such that it releases a certain amount of sugar into a chamber where it is mixed with soda. A random sample of 25 beverages was found to have a mean sugar content of $\overline{x} = 1.10$ and a stand deviation of $s = 0.015$. Find a 95% CI on the mean of sugar dispensed.

Sol We are given only $n = 25$, $\overline{x} = 1.10$, $s = 0.015$, so $95 = (1-\alpha)100$, hence $\alpha = 0.05$. The interval is: $[\overline{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \overline{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}] = [1.10 - t_{0.025,24}\frac{0.015}{\sqrt{25}}, 1.10 + t_{0.025,24}\frac{0.015}{\sqrt{25}}] = [1.10 - 2.064\frac{0.015}{\sqrt{25}}, 1.10 + 2.064\frac{0.015}{\sqrt{25}}] = [1.094, 1.106]$. We used table V, and moreover we assumed normal distribution (half of the exercises ...)!

As in the previous lecture: the one sided $100(1-\alpha)\%$ confidence bounds/intervals are defined via: $\overline{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}$ — lower-confidence bound,

$\overline{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}$ — upper-confidence bound;

Remember: t distribution IS used only in the case of normal population with unknown variance!

Do at the blackboard 8-24/page 272.

Last Example

Nine measurements of ozone concentration are obtained during a year:

3.5 5.1 6.6 6.0 4.2 4.4 5.3 5.6 4.4

Assuming the measurements are normally distributed, provide two 95% confidence intervals for the population mean μ:

- one assuming that the population variance $\sigma^2 = 1.21$;
- another *relaxing this assumption*.

First, $\alpha = 1 - 0.95$. Second, the sample mean is 5.01.

Assuming the population variance is 1.21:

$$5.01 \pm (1.96 \times \sqrt{1.21}/\sqrt{9}) \Rightarrow 5.01 \pm 0.72 \text{ or } (4.29, 5.73)\,.$$

If we relax the assumption, we have to compute

$$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 5.01 \pm 2.306\frac{0.97}{\sqrt{9}}.$$

$$= (4.26, 5.76),$$

where the pleasure of computing s belongs to ...

Summary of confidence intervals for μ when σ known

- If $\bar{x}$ is the size-n-sample mean from a normal population with known variance σ^2, the *exact* $(1-\alpha)$ C.I. for μ:

$$\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n};$$

- If $\bar{x}$ is the size-n-sample mean from a non-normal population with known variance σ^2 and n is 'big', the *approximate* $(1-\alpha)$ C.I. for μ:

$$\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n};$$

Summary of confidence intervals for μ when σ unknown

- If $\bar{x}$ is the size-n-sample mean from a normal population with unknown variance, the *exact* $(1-\alpha)$ C.I. for μ:

$$\bar{x} \pm t_{\alpha/2,n-1} s/\sqrt{n};$$

- If $\bar{x}$ is the size-n-sample mean from a normal or non-normal population with unknown variance and n is 'big', the *approximate* $(1-\alpha)$ C.I. is:

$$\bar{x} \pm z_{\alpha/2} s/\sqrt{n};$$

8-5 Estimating a population proportion, confidence interval

Consider the sample proportion $\hat{P} = X/n$, where X is the number of successes among the n trials. (For us: belonging to a class of interest.)

X follows a binomial distribution with parameters n and p. If n is large then X follows approximately a normal distribution. Since $\hat{P}$ is a linear function of X, then it is also approximately normal. Hence:

$\hat{P}$ follows $N(p, \frac{p(1-p)}{n})$ distribution approximately, i.e., $Z = \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is standard normal. Thus $1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}) = P(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}})$.

Therefore, for large n, a $100(1-\alpha)\%$ C.I. for the true proportion p is: $[\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}]$.

Problem: The confidence interval involves p, which is an unknown quantity. In practice, we simply replace p with its point estimate $\hat{p}$ to obtain an approximate confidence interval. For large n, that is: $n\hat{p} > 5$ and $n(1-\hat{p}) > 5$, a $100(1-\alpha)\%$ confidence interval for p is:

$$[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

One-sided confidence bounds are given by:

— lower bound: $\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p$;

— upper bound: $p \leq \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Do 8-47 at the blackboard! a) $n = 300$, $\hat{p} = \frac{13}{300}$, $100(1-\alpha) = 95$, so $\alpha = 0.05$, hence $\alpha/2 = 0.025$. The CI is $[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$ $= [0.04333 - 1.96\sqrt{\frac{0.04333(0.95667)}{300}}, 0.04333 + 1.96\sqrt{\frac{0.04333(0.95667)}{300}}] = [0.02029, 0.06637]$.

b) The CI is $(-\infty, \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}] = (-\infty, 0.04333 + z_{0.05}\sqrt{\frac{0.04333(0.95667)}{300}}] = (-\infty, 0.06273]$, since $z_{0.05} \cong 1.65$.

SAMPLE SIZE: If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error $|\hat{p} - p|$ will not exceed a specified amount E when the sample size is $n = (\frac{z_{\alpha/2}}{E})^2 \times p(1-p)$, convince yourself by plugging in the C.I.!!! BUT Problem: The latter formula involves the unknown parameter p. Using the maximization of $p(1-p)$ one gets:

RULE: If $\hat{p}$ is used as an estimate of p, we can be at least $100(1-\alpha)\%$ confident that the error $|\hat{p}-p|$ will not exceed a specified amount E when the sample size is $n = (\frac{z_{\alpha/2}}{E})^2 \times (0.25)$

Do 8-54/page 281. SOL: $99 = (1-\alpha)100$ implies that $\alpha = 0.01$, so$z_{0.005} \cong 2.58$, note that $E = 0.017$, so $n = (\frac{z_{\alpha/2}}{E})^2 \times (0.25) = 5758.13$, so we may take $n = 5759$.

8-4 Confidence interval for the variance

DEF: Consider a rv χ^2 whose p.d.f. is given by $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$ for $x > 0$. We say that χ^2 has a chi-square (χ^2) distribution with $k-1$ degrees of freedom. Properties:

1) the expectation is k, variance is $2k$;

2) NOT symmetric, in fact skewed to the right; bigger k implies more symmetric;

3) the percentile $P(\chi^2 > \chi^2_{\alpha,k}) = \alpha$ (an integral, right?) are found in table IV/714. For example: $\chi^2_{0.95,6} = 1.64$ and $\chi^2_{0.05,6} = 12.59$.

Theorem: Let $X_1, \ldots, X_n$ be a random sample from a NORMAL distribution with mean μ and variance σ^2, let S^2 be the sample variance.

Then the rv $\frac{(n-1)S^2}{\sigma^2}$ has a chi-square (χ^2) distribution with $n-1$ degrees of freedom.

Hence a $100(1-\alpha)$ CONFIDENCE interval for σ^2 is

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

. To get a C.I. for σ just take radicals!

ONE-SIDED bounds of variance:

— lower bound for σ^2 is $\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}}$, or the C.I. is $[\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}}, \infty)$

— upper bound for σ^2 is $\frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}$, or the C.I. is $(-\infty, \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}]$

Do exc 8-41/page 276 SOL: We get $n = 51$, $s = 0.37$, $100(1-\alpha) = 95$; so $\alpha/2 = 0.025$. We get first a C.I. for σ^2, and then for σ. We have $[\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}] = [\frac{(51-1)(0.37)^2}{\chi^2_{0.025,50}}, \frac{(51-1)(0.37)^2}{\chi^2_{0.975,50}}] = [\frac{(51-1)(0.37)^2}{71.42}, \frac{(51-1)(0.37)^2}{32.36}] = [0.096, 0.2115]$, so a C.I. for σ is $[\sqrt{0.096}, \sqrt{0.2115}]$, or $[0.31, 0.46]$.