Theory

Tuesday, January 7, 2025 4:01 PM

- Until now we used estimators to estimate an exact value for the unknown parameter
- However, it is more usefull (and reliable) to get an interval for this parameter. Such an interval is called confidence Interval
- Definition: Let $X_1, \dots, X_n$ sample of size n drawn from a distibution $F(x;\theta)$ with unknown parameter θ and a a small constant, $0<a<1$.
  Let $L = T_1(X_1, \dots, X_n)$ and $U = T_2(X_1, \dots, X_n)$ estimators so that:
  - $P(L \leq U) = 1$
  - $P(L \leq \theta \leq U) = 1-a \Leftrightarrow P\{L > \theta \text{ or } U < \theta\} = a$

  Then, the interval $[L, U]$ is the $100(1-a)\%$ confidence interval for the unknown parameter θ.
- Explanation: The $100(1-a)\%$ of the possible values for θ lies inside $[L, U]$
- Useful distribution

Distribution	Useful Properties
X^2: Αν $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$, τότε $X \sim X_n^2$	—
Student (t): $T_n = \dfrac{Z}{\sqrt{\frac{X_n^2}{n}}}$	$-t_{a,n} = t_{1-a,n}$
Γάμμα (Γ): $f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{a-1}}{} \\ 0 \end{cases}$	—
Fischer (F): $F_{n,m} = \dfrac{X_n^2/n}{X_m^2/m}$	$\dfrac{1}{F_{a,n,m}} = F_{1-a,m,n}$

- Other properties

Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. Then $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ and $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$

	Cases	Unknown Parameter	Useful Distribution	Two-sided CI	Single sided Lower CI	Single sided Upper CI
One population	Known σ^2	μ	Κανονική	$\bar{X} \pm z_{\frac{a}{2}} \frac{\sigma}{\sqrt{n}}$	$\left(-\infty, \bar{X} + z_a \frac{\sigma}{\sqrt{n}}\right)$	$\left(\bar{X} - z_a \frac{\sigma}{\sqrt{n}}, +\infty\right)$
	Unknown σ^2	μ	Student t	$\bar{X} \pm t_{\frac{a}{2}, n-1} \frac{S}{\sqrt{n}}$	$\left(-\infty, \bar{X} + t_{a,n-1} \cdot \frac{S}{\sqrt{n}}\right)$	$\left(\bar{X} - t_{a,n-1} \frac{S}{\sqrt{n}}, +\infty\right)$
	Unknown μ	σ^2	χ^2	$\left(\frac{(n-1)S^2}{\chi^2_{\frac{a}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{a}{2}, n-1}}\right)$	$\left(0, \frac{(n-1)S^2}{\chi^2_{1-a, n-1}}\right)$	$\left(\frac{(n-1)S^2}{\chi^2_{a, n-1}}, +\infty\right)$
Two populations	Known σ_X^2, σ_Y^2	$\mu_X - \mu_Y$	Κανονική	$\bar{X} - \bar{Y} \pm z_{\frac{a}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$	$\left(-\infty, \bar{X} - \bar{Y} + z_a \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$	$\left(\bar{X} - \bar{Y} - z_a \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, +\infty\right)$
	Unknown σ_X^2, σ_Y^2 but $\sigma_X^2 = \sigma_Y^2$	$\mu_X - \mu_Y$	Student	$\bar{X} - \bar{Y} \pm t_{\frac{a}{2}, n+m-2} \cdot \overset{*}{s_p} \sqrt{\frac{1}{n} + \frac{1}{m}}$	$\left(-\infty, \bar{X} - \bar{Y} + t_{a, n+m-2} \cdot \overset{*}{s_p} \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$	$\left(\bar{X} - \bar{Y} - t_{a, n+m-2} \cdot \overset{*}{s_p} \sqrt{\frac{1}{n} + \frac{1}{m}}, +\infty\right)$

$$* \quad s_p = \sqrt{\frac{(n-1)\, s_X^2 + (m-1)\, s_Y^2}{n+m-2}}$$

Exercise 1

Tuesday, January 7, 2025 4:27 PM

To study the time to live of a device produced in a factory, the following sample was drawn (in hours):

167, 184, 165, 174, 167, 180, 168, 173, 162

If the time to live of the device is normally distributed with variance 219.04, calculate the 95% confidence interval for the mean time to live of the devices produced in this factory.

It is given that $Z_{0.025}$=1.96, $Z_{0.05}$=1.64, $Z_{0.1}$=1.28.

$\sigma^2 = 219.04$

$(1-a)\cdot 100\% = 95\% \Rightarrow \boxed{a = 0.05}$

Known variance and we want to calculate the CI for the time to live of devices.

Thus $\Delta E = \bar{X} \pm Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$$\bar{X} = \frac{\sum_{i=1}^{9} x_i}{n} = \frac{167+184+165+174+167+180+168+173+162}{9} \Rightarrow \boxed{\bar{x} \simeq 171.11}$$

$$Z_{\frac{a}{2}} = Z_{\frac{0.05}{2}} = Z_{0.025} = 1.96$$

$$Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{\sqrt{219.04}}{\sqrt{9}} = 1.96 \cdot \frac{14.8}{3} \Rightarrow \boxed{Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \simeq 9.67}$$

$$CI = [171.11 - 9.67, 171.11 + 9.67] \Rightarrow \boxed{CI \simeq [161.44, 180.78]}$$

Exercise 2

Tuesday, January 7, 2025 4:27 PM

In a study concerning the concentration of a substance in cow's milk, a sample from 20 different cows was drawn. The concentration (in ppm) found in each of these cows is:

16, 0, 0, 2, 3, 6, 8, 2, 5, 0, 12, 10, 5, 7, 2, 3, 8, 17, 9, 1

Calculate the 95% confidence interval for the mean concentration in this substance.

$(1-a)\cdot 100\% = 95\% \Rightarrow \boxed{a=0.05}$

$n=20$

Unknown variance and we want to calculate the CI for the mean concentration. Thus:

$CI = \bar{x} \pm t_{\frac{a}{2},n} \cdot \frac{s}{\sqrt{n}}$

$$\bar{x} = \frac{16+0+0+2+3+6+8+2+5+0+12+10+5+7+2+3+8+17+9+1}{20} \simeq 5.8$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{20} x_i^2 - n\bar{x}^2\right)$$

16	0	0	2	3	6	8	2	5	0	12	10	5	7	2	3	8	17	9	1
256	0	0	4	9	36	64	4	25	0	144	100	25	49	4	9	64	289	81	1

$$\sum_{i=1}^{20} x_i^2 = 1164$$

$$s^2 = \frac{1}{n-1}\left(1164 - 20\cdot 5.8^2\right) \Rightarrow s^2 = 25.9 \Rightarrow \boxed{s \simeq 5.085}$$

$$CI = \left[5.8 - t_{\frac{0.05}{2},20}\cdot\frac{5.085}{\sqrt{20}},\ 5.8 + t_{\frac{0.05}{2},20}\cdot\frac{5.085}{\sqrt{20}}\right] =$$

$$= \left[5.8 - t_{0.025,20}\cdot\frac{5.085}{\sqrt{20}},\ 5.8 + t_{0.025,20}\cdot\frac{5.085}{\sqrt{20}}\right] =$$

$$= \left[5.8 - 2.093\cdot\frac{5.085}{\sqrt{20}},\ 5.8 + 2.093\cdot\frac{5.085}{\sqrt{20}}\right] \Rightarrow$$

$$\Rightarrow \boxed{CI \simeq [3.42, 8.18]}$$

Exercise 3

Tuesday, January 7, 2025 4:27 PM

Studies have shown that the weight of salmon raised in a fish farm follows a normal distribution with a mean value that varies from year to year but with a constant standard deviation of 0.3 kg. If we want to be 90% sure that our estimate of the average weight of salmon is correct, with a maximum error of ±0.1 kg, how large does the sample we take need to be in order to make this estimate using a confidence interval?

$\sigma = 0.3$

$(1-a)100\% = 90\% \Rightarrow \boxed{a = 0.1}$

$$\left.\begin{aligned} L &= \bar{x} - z_{\frac{a}{2}} \frac{\sigma}{\sqrt{n}} \\ U &= \bar{x} + z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \end{aligned}\right\} \Rightarrow U - L = \cancel{\bar{x}} + z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \cancel{-\bar{x}} + z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{U - L = 2 \cdot z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$$

$$U - L = 0.1 - (-0.1) = 0.2 \Rightarrow$$

$$\Rightarrow 2 \cdot z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq 0.2 \Rightarrow \boxed{z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq 0.1} \quad (1)$$

$$(1) \Rightarrow 1.64 \cdot \frac{0.3}{\sqrt{n}} \leq 0.1 \Rightarrow \sqrt{n} \geq 1.64 \cdot \frac{0.3}{0.1} = 4.92 \Rightarrow$$

$$\left.\begin{aligned} &\Rightarrow \boxed{n \geq 24.2} \\ &n \in \mathbb{N} \end{aligned}\right\} \Rightarrow$$

$\Rightarrow$ The smallest sample size we need in order to be 90% sure is 25

Exercise 4

Tuesday, January 7, 2025 4:27 PM

The following data are grades of male students in a school:

3, 8, 4, 11, 8, 6, 9, 10, 5

For the same test, we have the following grades of female student in the school:

16, 13, 20, 16, 15, 13

Find a 95% confidence interval for the difference $\mu_1 - \mu_2$ in the mean grade of male and female students, assuming that the variance is common and that the grades in both cases come from a normal distribution (although in reality they take discrete values).

$(1-a)\cdot 100\% = 95\% \Rightarrow \boxed{a = 0.05}$

Unknown but equal variances:

Άρα $\Delta E = \bar{x} - \bar{y} \pm t_{\frac{a}{2}, n+m-2} \cdot s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}$

$\bar{x} = \dfrac{3+8+4+11+8+6+9+10+5}{9} \Rightarrow \boxed{\bar{x} \simeq 7.11}$

$\bar{y} = \dfrac{16+13+20+16+15+13}{6} \Rightarrow \boxed{\bar{y} \simeq 15.5}$

$s_x^2 = 7.61$

$s_y^2 = 6.7$

$s_p^2 = \dfrac{(n-1)\cdot s_x^2 + (m-1)\cdot s_y^2}{n+m-2} = \dfrac{(9-1)\cdot 7.61^2 + (6-1)\cdot 6.7^2}{9+6-2} \simeq 7.26 \Rightarrow$

$s_p \simeq \sqrt{7.62} \Rightarrow \boxed{s_p \simeq 2.7}$

$t_{\frac{a}{2}, n+m-2} \cdot s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} = t_{0.025, 13} \cdot 2.7 \cdot \sqrt{\frac{1}{9} + \frac{1}{6}} \simeq 3.07$

$CI = [7.11 - 15.5 - 3.07,\ 7.11 - 15.5 + 3.07] \Rightarrow \boxed{CI = [-11.46, -5.32]}$

Exercise 5

Thursday, January 8, 2026 4:17 PM

Independent random samples are taken from the production of two machines on a production line. The weight of each item is under consideration. From the first machine, a sample of size 36 is taken, with a mean sample weight of 120 grams and a sample standard deviation of 4. A sample of size 64 is taken from the second machine, with a mean sample weight of 130 grams and a sample variance of 5. It is assumed that the weights of the items from the first machine are normally distributed with a mean value of μ1 and that the weights from the second machine are normally distributed with a mean value of μ2. Find a 99% confidence interval for the difference between the populations, μ1 - μ2:

a. if the population variances are unknown but equal.
b. if the population variances are 4 and 5, respectively.

$99\% = (1-a)100\% \Rightarrow \boxed{a = 0.01}$

$n = 36$ $\quad$ $m = 64$

$\bar{x} = 120$ $\quad$ $\bar{y} = 130$

$s_x = 4$ $\quad$ $s_y^2 = 5$

a) Unknown, but equal variances

$$CI = \bar{x} - \bar{y} \pm t_{\frac{a}{2}, n+m-2} \cdot s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$s_p^2 = \frac{(n-1) \cdot s_x^2 + (m-1) \cdot s_y^2}{n+m-2} = \frac{(36-1) \cdot 16 + (64-1) \cdot 5}{36+64-2} \simeq 8.93 \Rightarrow \boxed{s_p \simeq 2.99}$$

$$t_{\frac{a}{2}, n+m-2} \cdot s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} = t_{0.005, 98} \cdot 2.99 \cdot \sqrt{\frac{1}{36} + \frac{1}{64}} = 2.63 \cdot 2.99 \cdot 0.21 \simeq 1.64$$

$$CI = [120 - 130 - 1.64, 120 - 130 + 1.64] \simeq [-11.64, -8.36]$$

b) $\sigma_x^2 = 4$

$\sigma_y^2 = 5$

Known variances, thus:

known variances, thus:

$$CI = \bar{x} - \bar{y} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

$$z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = z_{0.005} \cdot \sqrt{\frac{16}{36} + \frac{25}{64}} = 2.56 \cdot 0.435 \simeq 1.11$$

$$CI = [120 - 130 - 1.11,\ 120 - 130 + 1.11] \Rightarrow \boxed{CI = [-11.11,\ -8.89]}$$