STAT-1302; Lecture 2; Jan. 11,'24

Last lecture was a review of CI's for μ when σ is known (§ 8.1, 8.2).

Ex. A city planner wants to estimate the average monthly residential water usage in the city. He selected a random sample of 39 households from the city, which gave the mean water usage to be 3,411.7 gallons over a one month period. Based on earlier data, the population standard deviation of the monthly residential water usage in this city is 387.5 gallons. Construct a 97% confidence interval for the average monthly residential water usage for all households in this city.

Sol'n:

Given: X = water usage; $\sigma = 387.5$

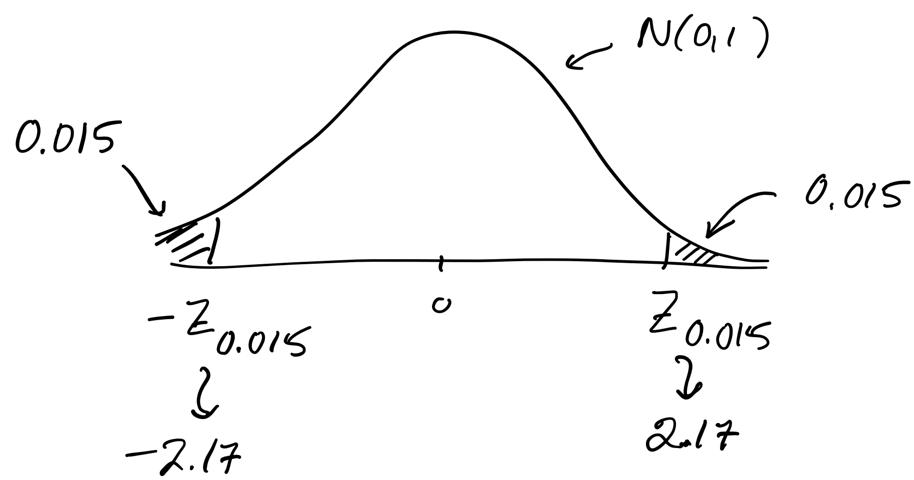
$X_1, \ldots, X_{39}$ is a random Sample and

$n = 39 > 30$ (CASE II).

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 3411.7 \pm 2.17 \frac{387.5}{\sqrt{39}} = *$$

$100(1-\alpha)\% = 97\% \Rightarrow 1-\alpha = 0.97, \ \alpha = 0.03$

$\alpha/2 = 0.015$



$* = 3,411.7 \pm 134.6534 = (3,277.0 , 3,546.4)$

Interpretation: We are 97% Confident that the true mean water usage in this City varies from 3,277.0 gallons to 3,546.4 gallons.

§ 8.3 CI for μ when σ is unknown.

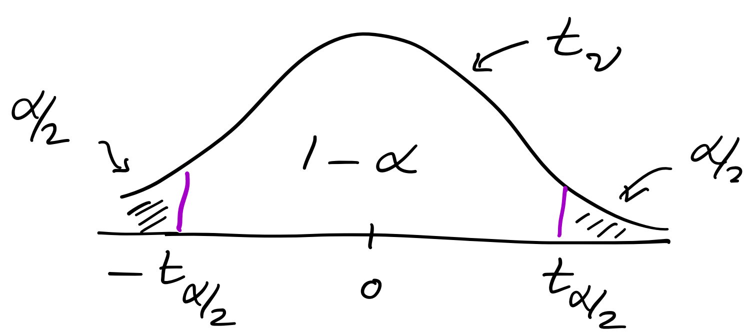
Recall: Let $X_1, \dots, X_n$ be a random Sample from a population with mean μ. Then,

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

t_{n-1}: a t-distribution with n-1 degrees of Parameter.

S = Sample standard deviation

Recall: $t_{\alpha/2} = ?$



$$1 - \alpha = P(-t_{\alpha/2} < t < t_{\alpha/2}) \qquad t \text{ r.v.}$$

$$= P\left(-t_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right)$$

$$= P\left(\bar{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$$

A 100(1−α)% CI for μ when σ is unknown is

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \qquad (d.f. = n-1).$$

$\boxed{\sigma \text{ unknown}}$

1. normal pop. 2. $n < 30$	$n \geq 30$	1. pop. not normal 2. $n < 30$
CASE I	CASE II	CASE III

CASE I, CASE II: Use $\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ as the CI for μ; σ unknown

CASE III: Use nonparametric Statistics.

Ex. A new alloy has been devised for use in a space vehicle. Tensile strength measurements are made on 15 pieces of alloy and the mean and standard deviation of these measurements are found to be 39.3 and 2.6, respectively.

Construct a 90% confidence interval for the mean tensile strength of this alloy. State any assumptions you are making.

Given:

X = tensile strength

$X_1, \ldots, X_{15}$ is a random sample ; $n = 15$;

$\bar{x} = 39.3$; $s = 2.6$.

Want: A 90% CI for μ.

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 39.3 \pm 1.761 \times \frac{2.6}{\sqrt{15}} = \times\times$$

$100(1-\alpha)\% = 90\%$; $1-\alpha = 0.9$ and $\alpha = 0.1$

$\therefore \alpha/2 = 0.05$.

$d.f. = n - 1 = 15 - 1 = 14$.

$t_{14;0.05} = ?$

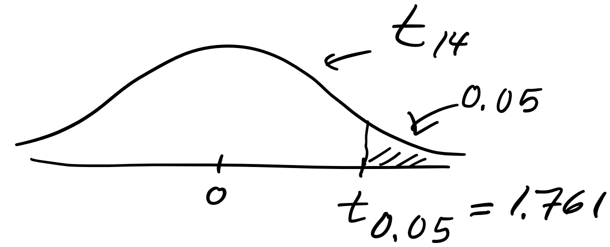


Table V

df	0.05
⋮	↓
14	$\boxed{1.761}$

$= (38.12, 40.48)$.

Since $n = 15 < 30$, we need X = tensile strength to be normally distributed.

Interpretation: We are 90% Confident that the true mean tensile strength is between 38.12 and 40.48.

Ex. In a lake pollution Study, the Concentration of lead in the upper Sedimentary layer of a lake bottom is measured from 25 Sediment Samples of 1,000 cm^3 each. The Sample mean and sample standard deviation of the measurements are found to be 0.36 and 0.06, respectively. Compute a 99% confidence interval for the mean concentration of lead per 1,000 cm^3 of sediment in the lake bottom. State any assumptions you are making.

Sol'n:

Given: X = lead concentration per 1,000 cm^3.

$X_1, \ldots, X_{25}$ is a random sample.

σ is unknown.

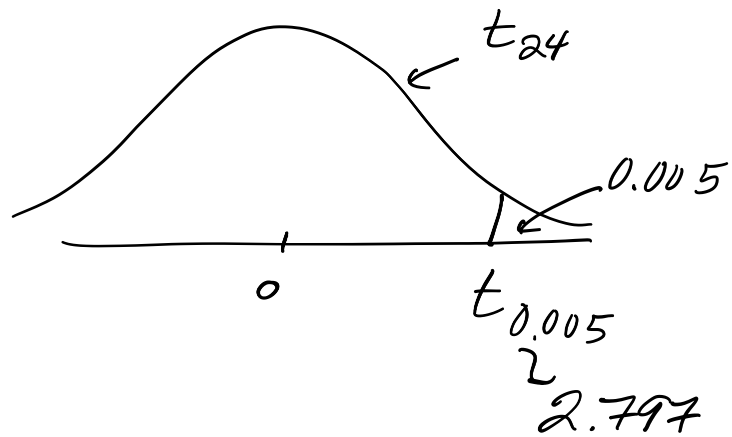
Assumption: (CASE I): X is normally distributed.

$$\bar{X} \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} = 0.36 \pm 2.797 \frac{0.06}{\sqrt{25}} = (***)$$

$$d.f. = n - 1 = 25 - 1 = 24.$$

$100(1-\alpha)\% = 99\%$, $1-\alpha = 0.99$; $\alpha = 0.01$

$\alpha/2 = 0.005$



$(\ast\ast\alpha) = (0.33,\ 0.39)$.

Interpretation: We are 99% Confident that the mean lead Concentration is between 0.33 and 0.39 per 1,000 cm^3.

($\bar{x}$ = Sample mean

S = Sample Standard deviation

μ = population mean

σ = population Standard deviation)

End of § 8.3.

§ 8.4 CI's for the Population Proportion (in large samples).

$$\hat{p} = \frac{\text{\# of items in a random sample with the characteristic of interest}}{n}$$

n = Sample size

$\hat{p}$ is a point estimator of the population proportion, p.

A $100(1-\alpha)\%$ CI for p (when $\frac{n}{N} < 0.05$ $n\hat{p} > 5$ and $n\hat{q} > 5$) is

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$\hat{q} = 1 - \hat{p}$.

Ex. According to a 2005 survey of 1,506 adult Americans, 75% of them said they frequently have sleep problems. Compute a 99% confidence interval for the percentage of all adult Americans who frequently have sleep problems.

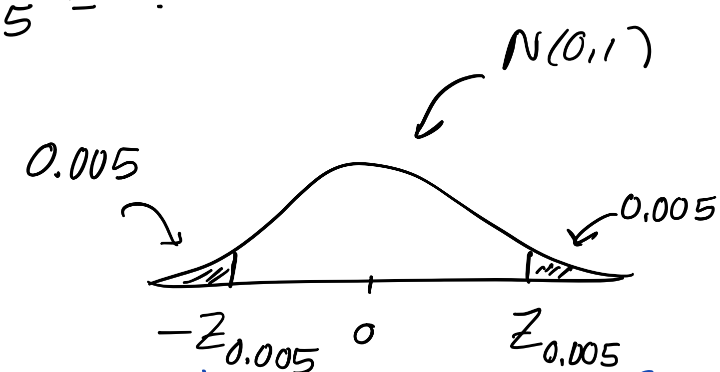
Sol'n:

Given: $n = 1506$, $\hat{p} = 0.75$

Show $\left\{ \begin{array}{l} n\hat{p} = 1506 \times 0.75 = 1129.5 > 5 \quad \text{and} \\ n\hat{q} = 1506 \times 0.25 = 376.5 > 5. \end{array} \right.$

$1 - \alpha = 0.99$; $\alpha = 0.01$; $\alpha/2 = 0.005$

$Z_{0.005} = ?$



-2.57 or -2.58

(Use Table IV.)

Will accept either 2.57 or 2.58

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.75 \pm 2.58 \sqrt{\frac{0.75 \times (1 - 0.75)}{1506}}$$

$$= 0.75 \pm 0.029$$

$$= (0.721 , 0.779).$$

Interpretation: We are 99% Confident that the percentage of adult Americans with sleep problems is between 72.1% and 77.9% ∎