Esercitazione 7 - Uni and Bilinear, Exact and Asymptotic Confidence Intervals

- $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known $\Rightarrow C_{1-\alpha}\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

Exercise 7.1

X = compressive strength

$X \sim N(\mu, \sigma^2)$ $\quad \sigma^2 = 1000$

$n = 12$ $\quad \bar{x}_{12} = 3255,42$

a) $CI_{.95}(\mu)$? $CI_{0,99}(\mu)$? compare widths

$X_1, \dots, X_{12} \overset{iid}{\sim} N(\mu, 1000)$ (σ^2 KNOWN)

$\Rightarrow CI_{1-\alpha}(\mu) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

$\bar{x}_n = 3255,42$

$\alpha = 0,05 \Rightarrow z_{.975} = 1,96 \Rightarrow CI_{.95}(\mu) = (3237.53, 3273.31$

$\alpha = 0,01 \Rightarrow z_{0,995} = 2,58 \Rightarrow CI_{.995}(\mu) = (3231.87, 3278.97)$

Lengths:

$\ell = 2 z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$\alpha = 0,05 \Rightarrow \ell = 35.78$

$\alpha = 0,01 \Rightarrow \ell = 47.10$

ℓ for lower α is wider.

b) One-sided 95% & 99% upper $CI(\mu)$

$$\left(-\infty, \bar{x}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

↳ since one-sided and we are only imposing the upper bound

just α, since we are not splitting the confidence of it being an outlier

$\alpha = 0{,}05 \Rightarrow z_{0,95} = 1{,}645 \Rightarrow CI_{0,95}(\mu) = (-\infty, 3270{,}44)$

$\alpha = 0{,}01 \Rightarrow z_{0,99} = 2{,}33 \Rightarrow CI_{0,99}(\mu) = (-\infty, 3276{,}69$

The lower α, the longer the interval.

A higher level random interval will contain μ with a higher probability, so the confidence interval will be longer.

c) $\left(\bar{x}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$ $\left(N, \sigma^2 \text{ known}\right)$

$\alpha = 0{,}05 \Rightarrow CI_{0,95}(\mu) = (3240{,}4, \infty)$

$\alpha = 0{,}01 \Rightarrow CI_{0,99}(\mu) = (3234.15, \infty)$

7.2

$n = 20$: measurements

$\bar{x}_n = 1{,}23$ $s^2 = 0{,}4$

↳ sample variance

a) $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $CI_{0,95}(\mu)$?

x : concentration in blood

$X_1, \dots, X_n \perp\!\!\!\perp$

$$\underbrace{CI_{1-\alpha}}_{\text{For unknown } \sigma^2} = \left(\bar{x}_n - t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}} \, , \, \bar{x}_n + t_{1-\frac{\alpha}{2}}(n-1)\frac{\sigma}{\sqrt{n}} \right)$$

$\bar{x}_{20} = 1{,}23$

$s = \sqrt{0{,}4}$

$n = 20$

$t_{0{,}975}(19) = 2{,}093$

$CI_{0{,}95}(\mu) = (0.934 \, , 1{,}526)$

b) What if σ^2 is known and equal to $\sigma^2 = 0{,}4$

$\sigma^2 = 0{,}4$

CI will be different since we use $z_{1-\frac{\alpha}{2}}$ rather than $t_{1-\frac{\alpha}{2}}$

$$CI_{1-\alpha}(\mu) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \, , \, \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

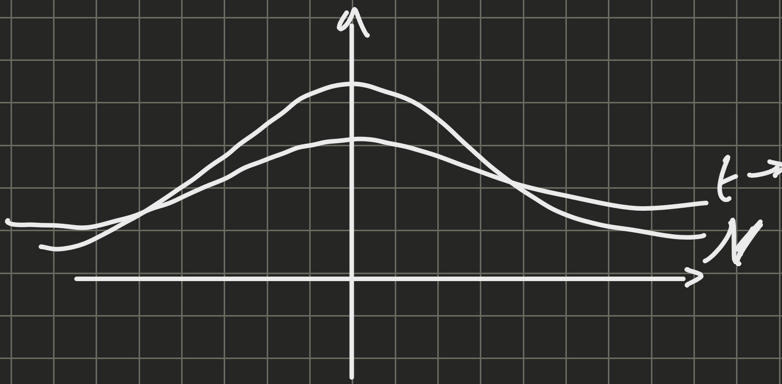
The quantiles are different

$\alpha = 0{,}05 \Rightarrow z_{0{,}975} \Rightarrow CI_{0{,}95}(\mu) = (0{,}953 , 1{,}507)$

c) In (b), we have a smaller interval because,

$t_{1-\frac{\alpha}{2}}(n-1) > z_{1-\frac{\alpha}{2}}$, for α small enough

(for us, this is always true)



7.3 d = distance to Vega from Earth · $\epsilon \sim N(0, \sigma^2)$, σ unknown

a) $D_1, \dots, D_{15} \overset{iid}{\sim} N(d, \sigma^2)$

$CI_{0,95}(d) = (25{,}09, 25.35)$ Symmetric

a) A point estimate for d, mean of population, is the sample mean

$$\bar{d}_n = \frac{25.09 + 25.35}{2} = 25{,}22$$

A point estimate for σ is the sample standard deviation s

$$s^2 = \frac{1}{n-1}\left(\left[\sum_i x_i^2\right] - n\bar{x}_n^2\right)$$

$$CI_{0,95}(d) = \left(\bar{d}_n - t_{0,975}(14)\frac{s}{\sqrt{15}}, \bar{d}_n + t_{0,95}(14)\frac{s}{\sqrt{15}}\right)$$

$$\overset{!}{=} (25{,}09, 25.35)$$

$$\bar{d}_n - t_{0,975}(14)\frac{s}{\sqrt{15}} = 25{,}09 \Rightarrow s = 0.235$$

b) At-most ⇒ calculate one-sided upper confidence bound of level 95%

At-least ⇒ same but lower.

$$\left(-\infty, \bar{d}_n + t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}}\right) = \left(-\infty, \bar{d}_n + t_{0,95}(14)\frac{s}{\sqrt{n}}\right)$$

$1-\alpha$, not $1-\frac{\alpha}{2}$ because one-sided

$= (-\infty, 25{.}3269)$

c) Same but for lowerbound

$$\left(\bar{d}_n - t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}, \infty\right) = (25.1131, \infty)$$

7.4

X: systolic blood pressure

$n = 25$, $\bar{x} = 120$ $\sum_{j=1}^{25} x_j^2 = 365400$

N sample with unknown μ and σ^2

a) Compute $CI_{0,95}$ one-sided upper confidence bound for μ.

$$CI_{0,95}(\mu) = \left(-\infty, \bar{x} + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}\right)$$

$\bar{x} = 120$

$t_{0,95}(24) = 1{,}7109$

$$S^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n} x_j^2 - n\bar{x}^2\right) = 225$$

$$\Rightarrow CI_{0,95}(\mu) = (-\infty, 125.1327)$$

b) $n = 39$ $\sum_{j=1}^{39} \tilde{x}_j^2 = 495885{,}8 \text{ mm}Hg^2$

NEW DATA

$\{x_1, \ldots, x_{64}\}$ → Combined new & old data

For the mean, we use the sample mean.

The new sample mean is a weighted average of the two sets

$$\bar{x}_{64} = \frac{25}{64} 120 + \frac{39}{64} 110 = 113.9062$$

$$S^2_{64} = \frac{1}{63}\left(\underbrace{\sum_{j=1}^{64} x_j^2}_{} - n\bar{x}^2_{64}\right) = 490.6344$$

$$\sum_1^{25} x_j^2 + \sum_{26}^{64} x_j^2 = 861285.8$$

c) $\mu > 109.28 \rightarrow$ how confident

The $(1-\alpha)100\%$ confidence lower bound is:

$$\left(\underbrace{\bar{x}_{64} - t_{1-\alpha}(63)\frac{s_{64}}{\sqrt{64}}}_{}, \infty\right)$$

set $=109.28$, then find α for which it is equal to that

We get:

$$t_{1-\alpha}(63) = 1.67$$

Being $n=64\ (>50)$

$\Rightarrow t_{1-\alpha} \simeq z_{1-\alpha}$ $\qquad$ t converges to N for $n \rightarrow \infty$

$\Rightarrow z_{1-\alpha} = 1.67 \Rightarrow 1-\alpha = 0.9525 \Rightarrow 95.25\%$

7.5

T: lifetime of electronic component

$t_1, \ldots, t_{100}$: $\sum_1^{100} t_i = 111.54 \quad \sum_1^{100} t_i^2 = 156.01$

a) An unbiased estimator for $\mathbb{E}(T)$ is the sample mean

$$\overline{T}_{100} = \frac{1}{100} \sum_{i=1}^{100} T$$

It's estimate is: $\frac{1}{100} \cdot 111.54 = 1.1154 := \bar{t}_n$

Estimator for variance is:

$$S_{100}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{100} T_i^2 - 100 \, \overline{T}_{100}^2 \right)$$

It's estimate is:

$$s^2 = 0{,}3192$$

b) Two-sided $CI_{0,95}(\mu)$?

We dont have gaussian assumption.

Since $n = 100 > 50$, $\Rightarrow$ large sample case $\Rightarrow$ CLT applicable

$$\Rightarrow CI_{1-\alpha}(\mu) = \left(\bar{t}_n - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} , \; \bar{t}_n + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right)$$

$\bar{t}_n = 1{,}1154$

$s = 0{,}56498$

$z_{0,975} = 1{,}96$

$$CI_{0,95}(\mu) = (1.0047, 1.2261)$$

c) How at most do you estimate μ at 95% confidence?

The asymptotic upper confidence bound of level 95% is:

$$\left(-\infty, \bar{t}_{100} + \underset{\sim}{z_{0,95}} \frac{s}{\sqrt{n}} \right) = (-\infty, 1.2083$$

1,645

7.6

$X \sim Poi(\lambda)$

$$\sum_{j=1}^{225} x_j = 202$$

a) Two-side $CI_{1-\alpha}(\lambda)$

$\hat{\lambda}$ is μ since $X \sim Poi(\lambda) \Rightarrow E(X) = \lambda$

$n = 225 \Rightarrow$ we can use large sample results

$$CI_{1-\alpha}(\lambda) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \; ; \; \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = (\underbrace{0.7740}_{t_1}, \underbrace{1.0216}_{t_2})$$

For Poi $E(X) = \lambda$

$Var(X) = \lambda \Rightarrow$ Plugin rule $\bar{x}$ is an estimator $Var(x)$

b) $CI_{0,95}(e^{-\lambda})$?

Since $g(\lambda) = e^{-\lambda}$ is strictly monotone (decreasing),

we can write $\mathbb{P}\{T_1 < \lambda < T_2\} = 0{,}95$

$$\Rightarrow P\{g(T_2) < \underbrace{g(\lambda)}_{=e^{-\lambda}} < g(T_1)\} = 0{,}95$$

Holds for monotonicity of g, we swapped since it's decreasing.

So, $CI_{0,95}(e^{-\lambda}) = (e^{-t_2}, e^{-t_1}) = (0{,}360, 0.461)$

Contime on 16/5

Exercise 7

$X, Y : E(X) = \theta \quad Var(X) = \theta(1-\theta) \quad \theta \in (0,1)$

$E(Y) = 1.1 \quad Var(Y) = 0.49 \quad Cov(X,Y) = 1$

$U = X + Y$

a) $E(U)$? $Var(U)$?

$E(U) = \theta + 1.1$

$Var(U) = Var(X) + Var(Y) + 2Cov(X,Y)$

$= \theta(1-\theta) + 0.49 + 2 = \theta(1-\theta) + 2.49$

b) $U_1, \dots, U_{100}$

is random sample

$\sum_{j=1}^{100} u_j = 160$

estimate $E(U)$, θ, $Var(U)$

An estimator for $E(U)$ is:

$\hat{U} = \bar{U}_{100} = \frac{1}{100} \sum_{j=1}^{100} U_j \rightarrow 1.6$ is the estimate

is sample mean

By plugin:

$\hat{\Theta} = \bar{U}_{100} - 1.1 \rightarrow 0{,}5$ is the estimate

$$Var(U) = \theta(1-\theta) + 2.49$$

By plugin:

$$\widehat{Var}(U) = \hat{\Theta}(1-\hat{\Theta}) + 2.49 \rightarrow 2.74 \text{ is the estimate.}$$

c) $\mu_U = E(U)$, $CI_{0.98}(\mu_U)$

$n=100 (\geq 50) \Rightarrow$ we can use CLT.

$$\left(\bar{U}_{100} - z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{\Theta}(1-\hat{\Theta})+2.49}}{\sqrt{n}}, \bar{U}_{100} + z_{1-\frac{\alpha}{2}} \frac{\sqrt{\dots}}{\sqrt{100}}\right) = C_{1-\alpha}(\mu_U)$$

Since we don't have sample variance, we use one estimate for the variance of U.

$\alpha = 0.1$ $CI(\mu_U) = (1.3277, 1.8723)$

d) $CI_{.98}(\theta)$

$\mu_U = \theta + 1.1 \rightarrow \theta = \mu_U + 1.1$

$CI_{0.90}(\theta) = (1.32277 - 1.1, 1.8723 - 1.1) = (0.2277, 0.7723)$

Exercise 8

p proportion of people (p: unknown), $\bar{x}_n$: sample mean.

a) Minimum n_0: $CI_{0.9}(p)$ has with < 0.05

X_i = response of the i-th person

$X_i \in \{0,1\}$, $X_i \sim Be(p) \Rightarrow E(X_i) = p$ $Var(X_i) = p(1-p)$

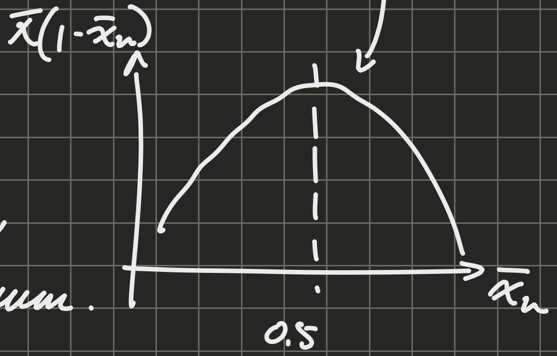
Assume $n > 50 \rightarrow$ so we can apply CLT

$$CI_{0,9}(p) = \left(\bar{x}_n - z_{.95}\frac{s}{\sqrt{n}}, \bar{x}_n + z_{0.95}\frac{s}{\sqrt{n}}\right)$$

$$\Rightarrow l = 2 z_{0.95} \frac{s}{\sqrt{n}} \leq 0.05 \rightarrow n_0 = 1083$$

We don't have s, but since $\hat{p} = \bar{x}_n \Rightarrow s = \sqrt{\bar{x}_n(1-\bar{x}_n)} = ?$
$\bar{x}_n$ is not known.

$\bar{x}_n = 0.5$ is the most conservative, since $\bar{x}_n$ is unknown, we use this to calculate the minimum.



b) 268 of n_0 are against the opening

Point estimate of p

$$\bar{x}_{n_0} = \frac{1}{n_0} \underbrace{\sum_{j=1}^{n_0} x_i}_{=268} = \frac{268}{1083} = 0{,}2475$$

c) $CI_{0.9}(p)$?

By asymptotic results:

$$\underset{\in(0,1)}{CI_{1-\alpha}(p)} = \left(\max\left\{0, \bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n(1-\bar{x}_n)}}{\sqrt{n_0}}\right\}, \min\left\{1, \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n(1-\bar{x}_n)}}{\sqrt{n_0}}\right\}\right)$$

since

$\alpha = 0{,}1$

$n_0 = 1083$

$\bar{x}_{n_0} = 0{,}2475$

$z_{0.95} = 1.645$

$\Rightarrow (0{,}2259 \,,\, 0{,}2691)$