改善

KAIZEN

TEAM

Ch 7

7-1 Point estimation

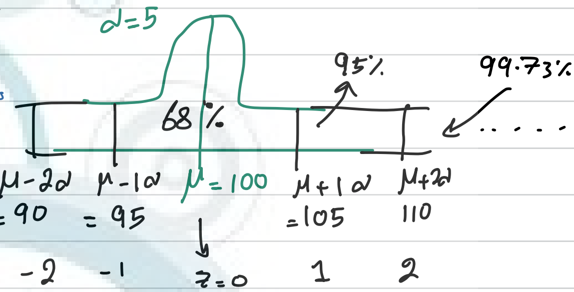
estimation → unknown

pop parameters → unknowns → estimation → from sample (sample statistic)

μ, σ^2, ?, π — Greek letters

Pop parameters: measure that are used to describe characteristic of the pop.

Central tendency →
- Mean — average — Sensitive to outlier
- median — value in the middle of values — not sensitive to outlier
- mode — most repeated



outlier → extreme value (0.27% may come out) * affects the range

variability →
- $Range = X_{max} - X_{min}$
- inter quartile range → $I.Q.R = Q_3 - Q_1$
- (σ^2) Variance $= \dfrac{\sum (X_i - \mu)^2}{N}$

n° central tendency (summation equation both sides)

$\sum (X_i - \mu) = 0$

$\sum_{i=1}^{n} X - \sum_{i=1}^{n} \mu = 0$

$\sum X_i - n\mu = 0$

$\dfrac{\sum X_i}{n} = \mu$

pop parameters → unknown → take sample → calc. sample statistic → estimate pop. parameter

Pop Parameter μ → $\bar{X}$ (point estimator) $\bar{x}$ (point estimate), $\sigma^2 \rightarrow S^2$ s^2, $\pi \rightarrow P$ $p^\wedge$

pop parameter θ → $\hat{\Theta}$ point estimator ← Hat: estimated

concerned with choosing a statistic that is a single # calculated from sample data for which we have some expectation that it is reasonably close to the parameter it is supposed to estimate iid (identical independent distribution)

$$\bar{X} = \frac{\sum x_i}{n}$$

$n=5 \rightarrow X_1, X_2, X_3, X_4, X_5$

$\hat{\mu}_1$: sample mean $\bar{X}$ $E(\bar{X}) = \mu$

$\hat{\mu}_2$: " median $\tilde{X}$: X_3 (R.V) $E(X_3) = \mu$

$\hat{\mu}_3 : \frac{1}{2}(X_1 + X_5)$ $E(\hat{\mu}_3) = \mu$

unbiased estimators

$\hat{\mu}_4 : X_1 + 0.5 X_5$ $E(\hat{\mu}_4) = 1.5\mu$

$\hat{\mu}_5 : \frac{1}{3}(X_2 + X_4)$ $E(\hat{\mu}_5) = \frac{2}{3}\mu$

biased estimators

Statistical properties :

unbiased estimator

$E(\hat{\Theta}) = \theta$, $\hat{\Theta}$ is unbiased estimator

$\neq \theta$, $\hat{\Theta}$ is biased estimator

$E(\hat{\Theta}) - \theta$ = biased $\rightarrow$ +ve over estimation

$\rightarrow$ -ve under estimation

Least variability → best

$V(\hat{\mu}_1) = V(\bar{X}) = \frac{\sigma^2}{5 \leftarrow n} = \frac{\sigma^2}{n}$ ← least variability

$V(\hat{\mu}_2) = \underset{\text{any point}}{V(X_3)} = \sigma^2$

$V(\hat{\mu}_3) = V\left(\frac{1}{2}(X_1 + X_5)\right) = \frac{1}{4}\left(V(X_1) + V(X_5)\right) = \frac{\sigma^2}{2}$

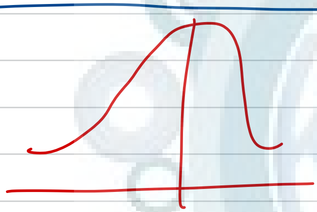
$\bar{X}$: minimum variance unbiased estimator (MVUE)

Minitab → population → $n = 25$

= 100 ↗ sample size

\# of sample

$\sigma = 5$



7-2 Sampling distribution & central limit theorem

$X \sim N(\underset{\text{mean}}{\mu}, \underset{\text{variance}}{\sigma^2})$

sample statistics

↓

Sampling distribution of the sample mean $\bar{X}$

minitab ex.

$X \sim N(100, 25)$

$\bar{X} \sim N(\underset{\mu}{100}, \underset{\frac{\sigma^2}{n}}{1})$

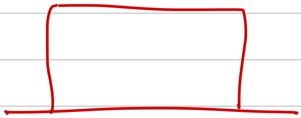
$p(\bar{X} > 110)$

$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \boxed{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}} = \frac{110 - 10}{5/\sqrt{25}} = 10$

Z =

$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}}$

,, $Z \sim N(\mu = 0, 1)$ standard normal



$X \sim U$

* sample distribution is always normal distribution $(\bar{X})$ whatever X distribution is

Central limit theorem: if $X_1, X_2, \dots, X_n$ is a random sample with size n from a population with mean (μ) & variance (σ^2) then the sampling distribution of the sample mean $\bar{X}$ will approximately follow normal dist. with mean (μ) & variance (σ^2/n)

* approximation depends on: 1 original pop. (distribution)
2 sample size ≥ 30

Standard error of an Estimator:

$$\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$$

$$MSE(\hat{\Theta}) = V(\hat{\Theta}) + (\text{bias})^2$$ "mean squared error"

Ch 8

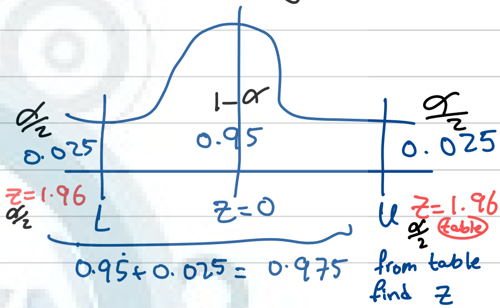
Statistical intervals for a single sample

How confident are you in your estimation?

8-1 Confidence intervals

Lets consider the random interval of values that should contain the true value of θ. A $100(1-\alpha)\%$ ($\alpha = 0.05$) confidence Interval (C.I) for a parameter θ, we should find a R.V whose expression involves θ & whose prob. distribution is at least approximately known

$$P(\underset{\text{lower}}{L} \le \theta \le \underset{\text{upper}}{U}) = 1-\alpha = 0.95$$



$$\hookrightarrow P(-1.96 \le z \le 1.96) = 0.95$$

$$P\left(-1.96 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 1.96\right) = 0.95$$

$$P\left(-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(1.96\frac{\sigma}{\sqrt{n}} \ge \mu-\bar{x} \ge -1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(\bar{x} - \underbrace{1.96}_{z_{\alpha/2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \underbrace{1.96}_{z_{\alpha/2}}\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

μ in interval → remains constant

- parameters → constant
- sample statistic → changes

$\bar{x}$ changes, $z_{\alpha/2}$, σ, $\sqrt{n}$

↳ remains constant

$\bar{X} = 100$ $\qquad$ $n \uparrow \quad C.I \downarrow$

#	α	$Z_{\alpha/2}$	n	σ	C.I
1	0.05	1.96	25	5	98.04 – 101.96
2	0.1	1.645	25	5	98.355 – 101.645
3	0.01	2.58	25	5	97.4 – 102.6
4	0.05	1.96	5	5	95.617 – 104.383
5	0.05	1.96	50	5	98.61 – 101.39
6	0.05	1.96	25	2.5	99.02 – 100.98
7	0.05	1.96	25	10	96.08 – 103.92

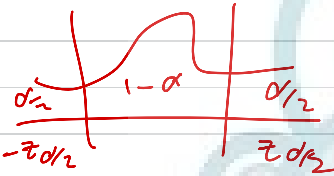
C.I for μ

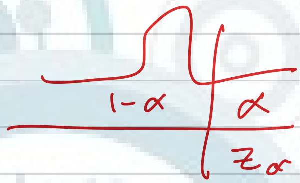
$$\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$ 2-sided - C.I (upper and lower)

upper $\left(\mu \leq \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}\right)$ 1-sided - C.I (upper or lower)

lower $\left(\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \mu\right)$

assuming σ^2 is known





pop. $\mu = ?$ $\sigma^2 = ?$ $\rightarrow$ $\bar{x}$, S^2

$$P(L < \mu < U) = 1-\alpha$$

$$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha$$

pop. $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ (R.V) $\rightarrow z \sim N(\mu = 0, \sigma^2 = 1)$ (standard) $\leftarrow$ σ^2 is known

sample $\dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ (R.V, R.V) higher variability than the other term $\leftarrow$ σ^2 is unknown

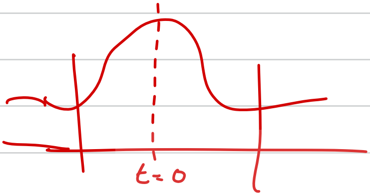
σ^2 is unknown if $n \geq 40 \rightarrow z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ S neglected

to use it: check if $\underline{X \sim N(\mu, \sigma^2)}$ pop. $\rightarrow$ normal dist.

$n < 40 \rightarrow t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ higher variability

t dist. (student dist.)

T-table



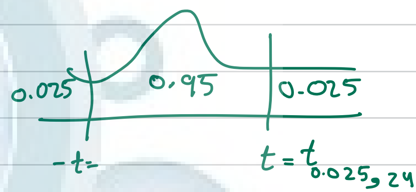
$P(t_{\alpha/2, n-1} < T < t_{\alpha/2, n-1}) = 1-\alpha$ $\Rightarrow$ switch $z \to t$, $\sigma \to S$

$\bar{X} = 100$, $n = 25$, $S = 5$, 95% C.I 2-sided

σ is unknown , S given , $n = 25 < 40 \to t$ dist.

$$\left(\bar{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}\right)$$

$$\left(100 - 2.064\frac{5}{\sqrt{25}} \le \mu \le 100 + 2.064\frac{5}{\sqrt{25}}\right)$$



$t = 2.064$

* σ is unknown , $n \geq 40$

↳ $\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$ 2-sided.

1-sided → upper / lower * z_{α} not $z_{\alpha/2}$

* σ is unknown , $n < 40$

$\left(\bar{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}\right)$ 2-sided

1-sided → upper / lower * t_{α} not $t_{\alpha/2}$

$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$ pop

sample

$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$ $n-1$: degrees of freedom (ν) nu

$S^2 = \frac{\sum x_i^2 - \frac{(\sum \bar{x}_i)^2}{n}}{n-1}$ calculation formula (easier)

$$\left.\begin{aligned} E(\bar{X}) &= \mu \\ E(S^2) &= \sigma^2 \end{aligned}\right\}$$ unbiased estimators

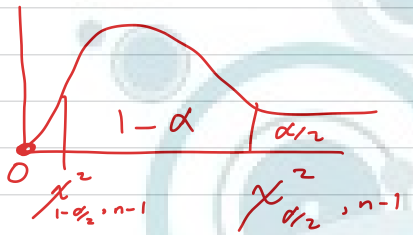
if $x_1, x_2 \ldots x_n$ is a random sample of size n with variance s^2 from a normal pop. with mean $= \mu$ & variance σ^2 then,

μ mean

s^2 variance

$$\frac{(n-1)\, s^2}{\sigma^2} \sim \chi^2 \left(n-1,\, 2(n-1)\right)$$

chi squared



$$P\left(\chi^2_{1-\alpha/2,\, n-1} \leq \chi^2 \leq \chi^2_{\alpha/2,\, n-1}\right) = 1-\alpha$$

$$\left(\quad < \frac{(n-1)s^2}{\sigma^2} < \quad \right)$$

$\vdots$

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,\, n-1}} < \sigma^2 < \frac{(n-1)\, s^2}{\chi^2_{1-\alpha/2,\, n-1}}\right) = 1-\alpha$$

Lower

Upper

$$P\left(\sqrt{\frac{(n-1)\, s^2}{\chi^2_{\alpha/2,\, n-1}}} < \sigma < \sqrt{\frac{(n-1)\, s^2}{\chi^2_{1-\alpha/2,\, n-1}}}\right)$$

ch 6

$(0 < \text{proportion} < 1)$

π pop. parameter $\longrightarrow$ p Sample statistic

proportion

X: # of males in sample

$$\hat{p} = \frac{X}{n}$$

C.I $\quad P(L < \pi < U) = 1 - \alpha$

$\pi \sim$ Binomial distribution

Binomial $\longrightarrow$ Normal approximation

symmetrical

if:

$$\left.\begin{aligned} n\cdot \overset{\pi}{p} &\geq 5 \\ n\cdot(1-\overset{\pi}{p}) &\geq 5 \end{aligned}\right\} \text{both}$$

$X = 16$, $n = 50$

$$\hat{p} = \frac{16}{50} = 0.32$$

$$P(L \leq \pi \leq U) = 1-\alpha$$

$$P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1-\alpha$$

$$z = \frac{x-\mu}{\sigma}$$

$$\mu = n \cdot \pi$$
$$\sigma^2 = n \cdot \pi(1-\pi)$$

$$z = \frac{x - n\cdot\pi}{\sqrt{n\cdot\pi(1-\pi)}} \quad \div n \rightarrow \frac{\frac{x}{n} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

$$z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1-\alpha$$

$$P\left(\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \pi \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$P\left(\quad \leq \pi \leq 0.32 + 1.96\sqrt{\frac{(0.32)(0.68)}{50}}\right)$$

0.191   0.449

Continous Data

$$\mu - \bar{x} = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

margin of error

$$n = \left(z_{\alpha/2} \frac{\sigma}{E} \right)^2$$

sample size

	$z_{\alpha/2}$	E	σ	n
95% C.I	1.96	10 cm	5	0.96 ~ 1
	1.96	5	5	3.8 ~ 4
	1.96	2.5	5	15.3 ~ 16
99% C.I	2.58	2.5	5	26.6 ~ 27
	1.96	2.5	10	61.4 ~ 62
	1.96	2.5	1	0.61 ~ 1

* always round **up**

Discrete data

$$\pi - \hat{p} = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$n = z_{\alpha/2}^{2} \frac{\hat{p}(1-\hat{p})}{E^2}$$

$$\frac{(1.96)^2 (0.32)(0.68)}{(0.1)^2} = 83.5 \sim 84$$

$$\bar{X} \pm E$$

$$C.I = 2E$$

dence interval with the one in part (a).

8.1.8 WP VS A civil engineer is analyzing the compressive strength of concrete. Compressive strength is normally distributed with $\sigma^2 = 1000(\text{psi})^2$. A random sample of 12 specimens has a mean compressive strength of $\bar{x} = 3250$ psi.

cont

a. Construct a 95% two-sided confidence interval on mean compressive strength.

b. Construct a 99% two-sided confidence interval on mean compressive strength. Compare the width of this confidence interval with the width of the one found in part (a).

a. $$P\left(3250 - 1.96\frac{\sqrt{1000}}{\sqrt{12}} \leq \mu \leq 3250 + 1.96\frac{\sqrt{1000}}{\sqrt{12}}\right)$$

$$P(3232.107 \leq \mu \leq 3267.892)$$

b. $$P\left(3250 - 2.58\frac{\sqrt{1000}}{\sqrt{12}} \leq \mu \leq 3250 + 2.58\frac{\sqrt{1000}}{\sqrt{12}}\right)$$

$$P(3226.448 \leq \mu \leq 3273.552)$$

8.1.9 WP Suppose that in Exercise 8.1.8 it is desired to estimate the compressive strength with an error that is less than 15 psi at 99% confidence. What sample size is required?

$$n = \left(z_{\alpha/2}\frac{\sigma}{E}\right)^2$$

$$= \left(2.58 \times \frac{\sqrt{1000}}{15}\right)^2 = 24.1 \simeq 25$$

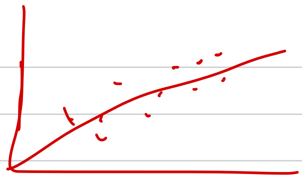
8.2.9 The compressive strength of concrete is being tested by a civil engineer who tests 12 specimens and obtains the following data:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295

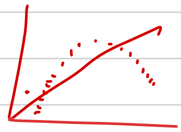
a. Check the assumption that compressive strength is normally distributed. Include a graphical display in your answer.

b. Construct a 95% two-sided confidence interval on the mean strength. T-table

c. Construct a 95% lower confidence bound (1- sided) on the mean strength. Compare this bound with the lower bound of the two-sided confidence interval and discuss why they are different.



→ Normal



→ not normal

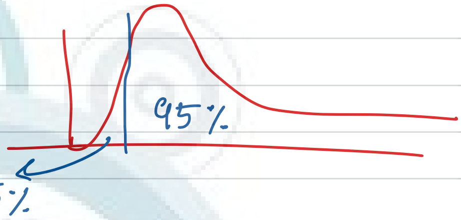
standard deviation of the skating performance test.

8.3.5 **WP** An article in *Technometrics* ["Two-Way Random Effects Analyses and Gauge R&R Studies" (1999, Vol. 41(3), pp. 202–211)] studied the capability of a gauge by measuring the weight of paper. The data for repeated measurements of one sheet of paper are in the following table. Construct a 95% one-sided upper confidence interval for the standard deviation of these measurements. Check the assumption of normality of the data and comment on the assumptions for the confidence interval.

Observations

3.481	3.448	3.485	3.475	3.472
3.477	3.472	3.464	3.472	3.470
3.470	3.470	3.477	3.473	3.474

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



$$\le \frac{(14)\,s^2}{6.57}$$

Bush.

8.4.2 **WP** **VS** An article in *Knee Surgery, Sports Traumatology, Arthroscopy* ["Arthroscopic Meniscal Repair with an Absorbable Screw: Results and Surgical Technique" (2005, Vol. 13, pp. 273–279)] showed that only 25 out of 37 tears (67.6%) located between 3 and 6 mm from the meniscus rim were healed.

a. Calculate a two-sided 95% confidence interval on the proportion of such tears that will heal.

b. Calculate a 95% lower confidence bound on the proportion of such tears that will heal.

$n = 37$ $\hat{p} = 0.676$

$x = 25$