Estimating a Sample Mean, True σ known

X : unknown μ
known σ

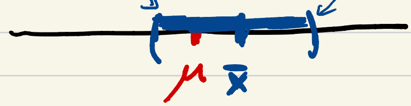
n samples

$\bar{x}$: sample mean

* we may be given a sample sd s but this is less information than the true sd σ
Ignore s and use z critical value.

2-sided $100(1-\alpha)\%$ Confidence Interval (CI)

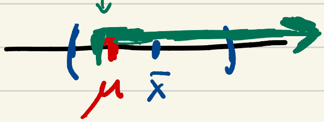
$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \mu \le \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$



1-sided $100(1-\alpha)\%$ CIs

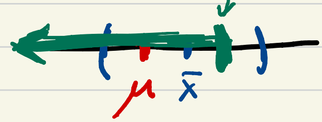
Upper CI
$\Leftrightarrow$ Lower Bound

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha} \le \mu$$



Lower CI
$\Leftrightarrow$ Upper bound

$$\mu \le \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha}$$



Estimating a Sample Mean, True σ unknown (and X normal)

X : unknown μ
unknown σ

$X \sim N(\mu, \sigma)$

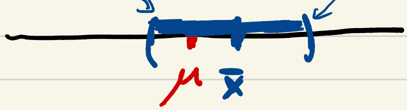
n samples

$\bar{x}$: sample mean

s : sample sd

2-sided $100(1-\alpha)\%$ Confidence Interval (CI)

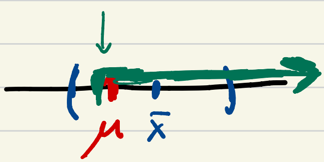
$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2} \le \mu \le \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2}$$



1-sided $100(1-\alpha)\%$ CIs

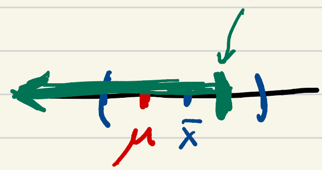
Upper CI
$\Leftrightarrow$ Lower Bound

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha} \le \mu$$



Lower CI
$\Leftrightarrow$ Upper bound

$$\mu \le \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha}$$



Estimating a Population Proportion, p

$X \sim \text{Binom}(n,p)$

n: num samples

p: true proportion of category 1.

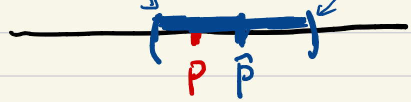
$\hat{p}$: $\frac{X}{n}$, proportion estimated in sample

SD of a binomial is known so we use z critical value and

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

2-sided 100(1-α)% Confidence Interval (CI)

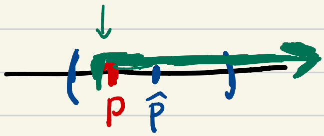
$$\hat{p} - \sigma_{\hat{p}} Z_{\alpha/2} \le p \le \hat{p} + \sigma_{\hat{p}} Z_{\alpha/2}$$



1-sided 100(1-α)% CIs

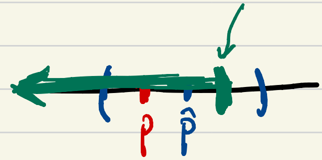
Upper CI ⟺ Lower Bound

$$\hat{p} - \sigma_{\hat{p}} Z_{\alpha} \le p$$



Lower CI ⟺ Upper bound

$$p \le \hat{p} + \sigma_{\hat{p}} Z_{\alpha}$$



Flowchart for two sample situations

Q1: Do we know the true st dev's σ_1, σ_2?

Yes: Use z critical value and

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

No

Q2: Are the observations in sample 1 and 2 paired in some way?

(Before / After ; Two meas from same person)

Yes: Use t critical value ($v = n-1$)

define $d_i = x_i - y_i$ and use $\bar{d}$ and s_d.

No?

Q3: Can we assume the true st. dev's are equal?

Yes: Use t critical value ($v = n_1 + n_2 - 2$)

pooled st dev: $$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

No: Use t critical value

St dev $$S_{12} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left((s_1^2/n_1)^2/(n_1-1)\right) + \left((s_2^2/n_2)^2/(n_2-1)\right)}$$ Rounded down

Estimating a Difference in Means (Two Samples)

Sample 1.
sample mean: $\bar{X}_1$
sample size: n_1

Sample 2
sample mean: $\bar{X}_2$
sample size: n_2

Sample 1 Variance: S_1^2

Sample 2 Variance: S_2^2

2-sided $100(1-\alpha)\%$ Confidence Intervals

* True SD's σ_1, σ_2 known

Define $\sigma_{12} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

$$(\bar{x}_1 - \bar{x}_2) - \sigma_{12} z_{\alpha/2} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + \sigma_{12} z_{\alpha/2}$$

* True SD's unknown, but assumed equal

Define pooled variance

$$S_p^2 = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1 + n_2 - 2}$$

$$(\bar{x}_1 - \bar{x}_2) - s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2}$$

$$\le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2}$$

Degrees of freedom: $v = n_1 + n_2 - 2$

Estimating a Difference in Means (Two Samples)

Sample 1.
sample mean: $\bar{X}_1$
sample size: n_1

Sample 2
sample mean: $\bar{X}_2$
sample size: n_2

Sample 1 Sample 2

$D_i = X_i - Y_i$

$\bar{D} = \frac{1}{n} \sum D_i$

$\mu_1 - \mu_2 = \mu_D$

2-sided $100(1-\alpha)\%$ Confidence Intervals

* True SD's σ_1, σ_2 unknown, unequal

Define $S_{12} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$$(\bar{x}_1 - \bar{x}_2) - S_{12}\, t_{\alpha/2} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + S_{12}\, t_{\alpha/2}$$

* see v below

* True SD's unknown, paired observations

Define s_d to be the sample st. dev. of the differences

$$\bar{d} - \frac{s_d}{\sqrt{n}} t_{\alpha/2} \le \mu_1 - \mu_2 \le \bar{d} + \frac{s_d}{\sqrt{n}} t_{\alpha/2}$$

Degrees of freedom: $v = n - 1$

* Satterthwaite approximation $\quad v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left((s_1^2/n_1)^2/(n_1-1)\right) + \left((s_2^2/n_2)^2/(n_2-1)\right)}$ Rounded down