Recall: $X_1, \dots, X_n$ iid normal $\sim N(\mu, \sigma^2)$.

Point estimate for $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$.

Confidence interval: $\mathbb{P}\left(\left| \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \right| < z_{\alpha/2} \right) = 1 - \alpha = 95\%$

$\alpha = 0.05$

If σ is known!

$z_{0.025} = 1.96$

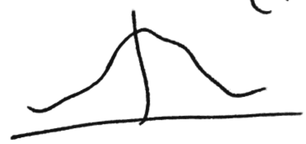
(*) $\mathbb{P}\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$

f_Z

If σ is unknown, estimate σ^2 by S^2.

$c_n \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}$.

$\mathbb{P}\left(\left| \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \right| < t_{\alpha/2, n-1} \right) = 1 - \alpha$

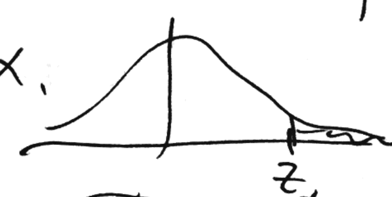


2.045

(**) $\mathbb{P}\left(\bar{X} - \quad < \mu < \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \right) = 1 - \alpha$.

95%

z_α means $\int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \alpha$.



~~Once $n \geq 30$, approx $t_{\alpha/2, n-1}$~~ by $z_{\alpha/2} = 1.96$.

Using z instead of t gives only 94% confidence

New setting: Which Brand is better?

Brand A: 40 bulbs, found avg lifespan 418 hrs "$= \bar{X}_A$"

Brand B: 50 bulbs, — — — 402 hrs. $\bar{X}_B$

Version I: assume "know" $\sigma_A = 26$, $\sigma_B = 22$.

Want to know $\mu_A - \mu_B$. 95% confidence interval.

$$(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) = (\bar{X}_A - \mu_A) - (\bar{X}_B - \mu_B).$$

Assume: lightbulb lifespans are iid Gaussian.

Gauss mean 0   Gauss mean 0 $\Rightarrow$ Gauss mean 0.

$\Rightarrow$ Gauss mean 0, $\mathrm{Var} = \dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}$.

$$\Rightarrow \mathbb{P}\left(\left| \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \right| < z_{\alpha/2} \right) = 1 - \alpha.$$

$\Rightarrow$ with $(1-\alpha)100\%$ confidence, $(\alpha = 0.05)$

$$\bar{X}_A - \bar{X}_B - z_{\alpha/2}\sqrt{\,''\,} < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}.$$

$\bar{X}_A = 418$, $\bar{X}_B = 402$, $z_{\alpha/2} = 1.96$, $\sigma_A = 26$, $\sigma_B = 22$, $n_A = 40$, $n_B = 50$

$-3 <$ < 6

(2)

More realistically, don't know σ!!!

$$(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) = \underbrace{(\bar{X}_A - \mu_A)} - \underbrace{(\bar{X}_B - \mu_B)}.$$

Try: $\dfrac{\bar{X}_A - \mu_A}{\sqrt{S_A^2/n_A}}$ = T distr with $n_A - 1$ degrees.

No simple ~~solution to~~ expression for distr of unless

$$\boxed{\sigma_A = \sigma_B = \sigma.}$$

$$\frac{\bar{X}_A - \mu_A}{\sqrt{\sigma_A^2/n_A}} \cdot \sqrt{\frac{\sigma_A^2}{\underbrace{(n_A-1)S_A^2}/(n_A-1)}}$$

$$\frac{1}{\sqrt{Y/\nu}}.$$

$\bar{X}_A - \mu_A - (\bar{X}_B - \mu_B) \leftarrow$ mean 0 Gaussian,

$$\text{Var} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \underline{\underline{\sigma^2}}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)$$

Obs: $\dfrac{(n_A-1)}{\sigma_A^2} S_A^2 + \dfrac{(n_B-1)}{\sigma_B^2} S_B^2 = \chi^2$ with $n_A + n_B - 2$ degrees of freedom.

$$\underline{\underline{\frac{1}{\sigma^2}}}\left[(n_A-1) S_A^2 + (\underline{n_B-1}) S_B^2 \right].$$

$$\Rightarrow \frac{\bar{X}_A - \mu_A - (\bar{X}_B - \mu_B)}{\sqrt{\cancel{\sigma^2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \cdot \sqrt{\frac{(n_A + n_B - 2)}{\frac{1}{\cancel{\sigma^2}}\left[(n_A - 1)S_A^2 + (n_B - 1)S_B^2\right]}}$$

$\Rightarrow$ <u>$1-\alpha$ confidence interval:</u>

$$\bar{X}_A - \bar{X}_B - \sim < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + t_{\alpha/2,\, n_A+n_B-2}\sqrt{\sim}$$

$$\sqrt{\underbrace{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}}_{\text{"pooled estimator"}}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

<u>E.g.</u>: 400 people got flu vaccine, 136 had reaction. Each $X_i = \begin{cases} 0 & \\ 1 & \text{had reaction} \end{cases}$ (Bernoulli).

$\mathbb{P}(X_i = 1) = \theta \leftarrow$ don't know. $\hat{\theta} = \bar{X} = \frac{136}{400}$.

$\theta = \mathbb{E}(X_i)$. What about $1-\alpha$ - confidence interval for θ?

$\bar{X} - \theta$

$$n\bar{X} = \cancel{\tfrac{1}{n}}\underbrace{\underbrace{\sum X_i}_{\text{Bernoulli}}}_{\text{binomial}}$$

<u>Recall</u>: $\mathbb{P}(n\bar{X} = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$.

$$\mathrm{Var}(X_i) = \theta(1-\theta) = \theta - \theta^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

$$\mathrm{Var}(\overline{X}) = \frac{\theta-\theta^2}{n}, \quad \mathrm{Var}(n\overline{X}) = n\theta(1-\theta).$$

$$\frac{n\overline{X} - n\theta}{\sqrt{n\theta(1-\theta)}} \approx N(0,1).$$

Binomial $\to$ Gaussian
(CLT)
DeMoivre-Laplace.

So $\underbrace{\text{approximately}}$ $1-\alpha$ confidence interval:

$$\mathbb{P}\left(\left| \frac{n(\overline{X}-\theta)}{\sqrt{n\theta(1-\theta)}} \right| < z_{\alpha/2} \right) \approx 1-\alpha.$$

Interval:

$$\underbrace{\left|\overline{X}-\theta\right| < z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}}$$

Not really, still have θ here

Tweak 1 (hwk): Solve algebra for interval in θ.

Lazier Tweak 2: Further Approximation: $\theta \approx \overline{X}$.

$$\left|(\overline{X}-\theta)\right| \lesssim z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}.$$

A