

SMGT 430: Don't be fooled by noise

$$\left. \begin{array}{l} \mu \sim \text{Normal}(\mu_0, \sigma_0^2) \\ \bar{Y} \sim \text{Normal}(\mu, \sigma^2/n) \end{array} \right\} \begin{array}{l} \mu \sim \text{Normal}(\mu_0, \sigma_0^2) \\ \bar{Y}_1 \sim \text{Normal}(\mu, \sigma^2/n_1) \\ \bar{Y}_2 \sim \text{Normal}(\mu, \sigma^2/n_2) \end{array} \text{ independent}$$

Which is a better predictor of \bar{Y}_2 : \bar{Y}_1 or μ_0 ? $\left. \begin{array}{l} \bar{Y}_1 = \mu + z_1 \quad z_1 \sim N(0, \sigma^2/n_1) \\ \bar{Y}_2 = \mu + z_2 \quad z_2 \sim N(0, \sigma^2/n_2) \end{array} \right\} \text{ ind.}$

$$E[(\bar{Y}_2 - \bar{Y}_1)^2] = \quad \text{vs.} \quad E[(\bar{Y}_2 - \mu_0)^2]$$

$$E[(\bar{Y}_2 - \mu + \mu - \bar{Y}_1)^2] =$$

$$E[(\bar{Y}_2 - \mu)^2 + 2(\bar{Y}_2 - \mu)(\mu - \bar{Y}_1) + (\mu - \bar{Y}_1)^2] =$$

$$\underbrace{E[(\bar{Y}_2 - \mu)^2]}_{\sigma^2/n_2} - 2 \underbrace{E[(\bar{Y}_2 - \mu)]}_{0} \underbrace{E[(\bar{Y}_1 - \mu)]}_{0} + \underbrace{E[(\bar{Y}_1 - \mu)^2]}_{\sigma^2/n_1} = \sigma^2/n_2 + \sigma^2/n_1$$

$$E[(\bar{Y}_2 - \mu_0)^2] =$$

$$E[(\bar{Y}_2 - \mu + \mu - \mu_0)^2] =$$

$$E[(\bar{Y}_2 - \mu)^2 + 2(\bar{Y}_2 - \mu)(\mu - \mu_0) + (\mu - \mu_0)^2] =$$

$$\underbrace{E[(\bar{Y}_2 - \mu)^2]}_{\sigma^2/n_2} + 2 \underbrace{E[(\bar{Y}_2 - \mu)]}_{0} \underbrace{E[(\mu - \mu_0)]}_{0} + \underbrace{E[(\mu - \mu_0)^2]}_{\sigma_0^2} = \sigma^2/n_2 + \sigma_0^2$$

σ^2/n_1 vs. σ_0^2 : equal when $n_1 = \sigma^2/\sigma_0^2$

Ex #1: True talent 3PT% is normal w/ mean 37%, s.d. 2%.
How many attempts before 1st half 3PT% is better predictor than league average?

$$\sigma^2 = p(1-p) = .37 \cdot .63 = .2331 \quad \sigma_0^2 = .02^2 = .0004 \quad n_1 = .2331 / .0004 = 583$$

What is the correlation between \bar{Y}_1 and \bar{Y}_2 ?

$$\text{Corr}(\bar{Y}_1, \bar{Y}_2) = \frac{\text{Cov}(\bar{Y}_1, \bar{Y}_2)}{\sqrt{\text{Var}(\bar{Y}_1)\text{Var}(\bar{Y}_2)}} = \frac{\sigma_0^2}{\sqrt{(\sigma_0^2 + \sigma^2/n_1)(\sigma_0^2 + \sigma^2/n_2)}} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n} \text{ if } n_1 = n_2 = n$$

Ex #2: What is correlation between 1st half and 2nd half 3PT% if every player has $n=400$ attempts in each half?

$$\sigma_0^2 / (\sigma_0^2 + \sigma^2/n) = .0004 / (.0004 + .2331/400) = .41$$