Motivation: optimal sample size allocation

Week 3 (3.4)

Stat 260, St. Clair

Tradeoff: Cost vs. Precision

As n (sample size) increases:

- SE's get decrease (more precise) but
- sampling costs increase

SRS example

- $N = 3000$ units
- Assume $S = 1$ for our measurement of interest

Cost: costs per unit is $c = \$2$

$$\text{total cost} = C(n) = \$2n$$

Precision: 95% margin of error for estimating the mean

$$ME(n) = 1.96 \times SE(\bar{y}_{srs}) = 1.96 \times \sqrt{\left(1 - \frac{n}{3000}\right)\frac{1}{n}}$$

SRS example: determine the n that...

$\$2 \cdot n$

$\min C(n) = \boxed{C(94)}$

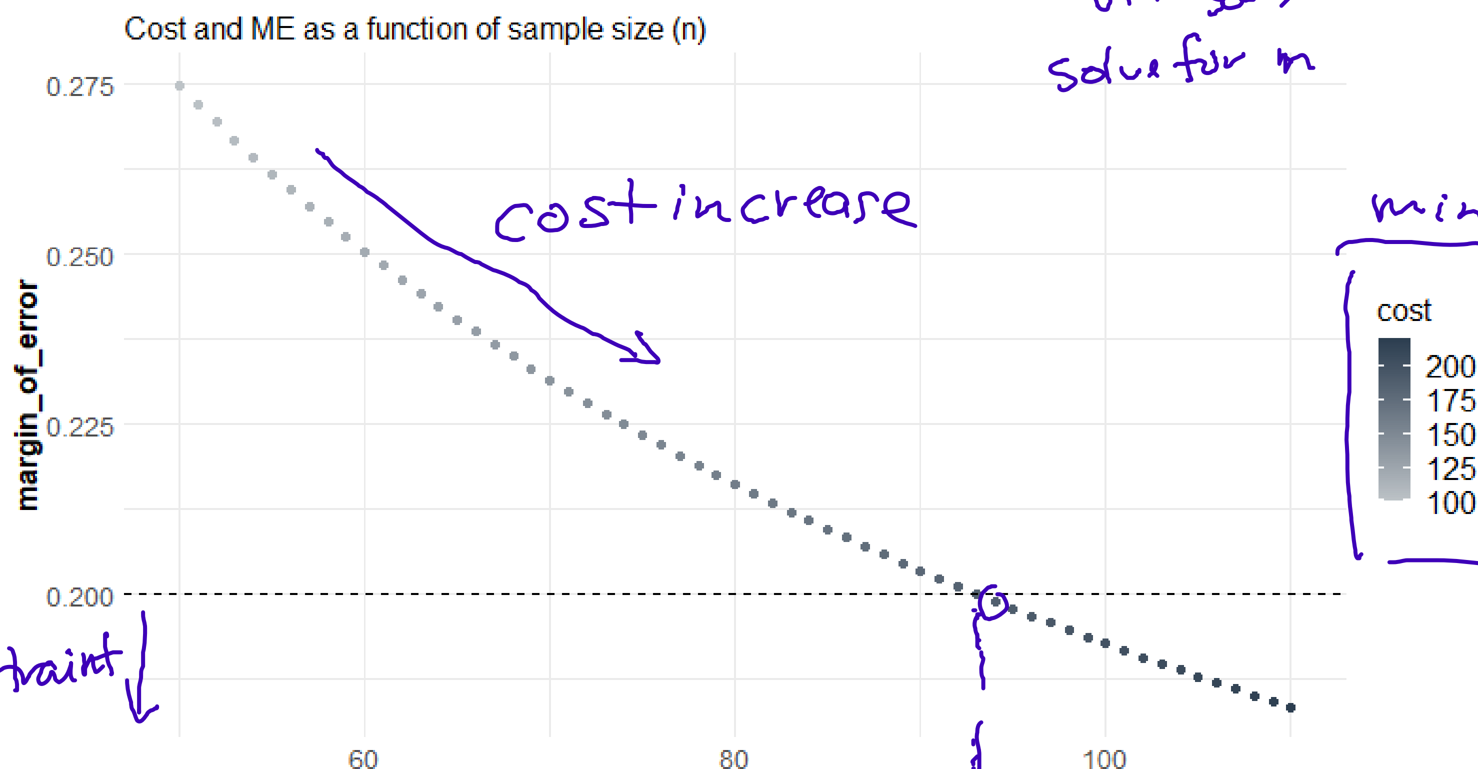
$\{ME(n) \le .2\}$

Constraint: ME of at most 0.2 $\Rightarrow \{n \geq 93.06\}$

Optimize: minimize cost under this constraint

$1.96\sqrt{\left(1-\frac{n}{3000}\right)\frac{1}{n}} \le .2$

solve for n



cost increase

min

constraint

$n \geq 93.06$

SRS example: determine the n that...

Constraint: Costs of at most \$200 $\Rightarrow n \le 100$

Optimize: minimize margin of error (SE) under this constraint

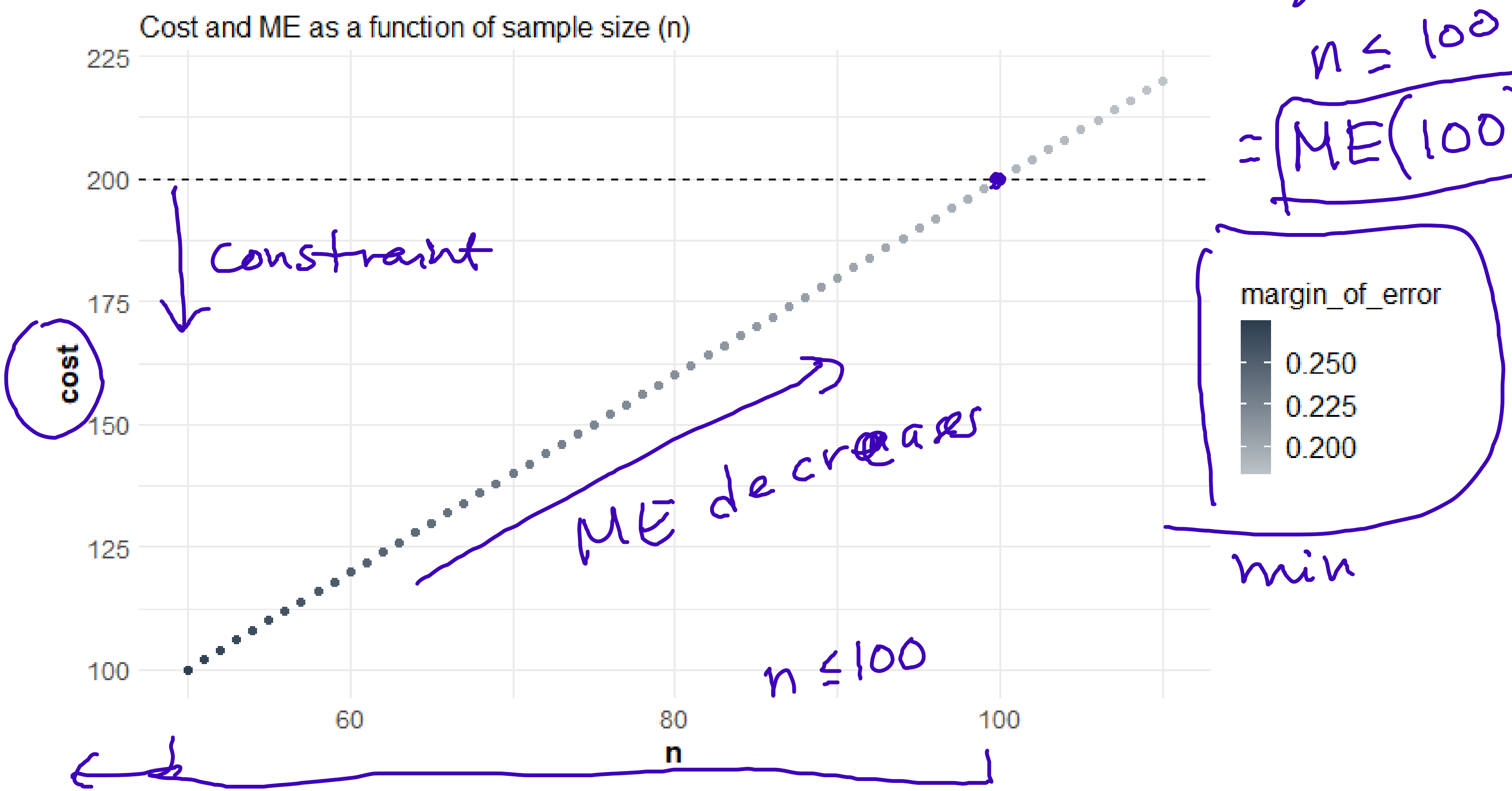
$\min ME(n)$

$\{C(n) \le \$200\}$

$\$2 \cdot n \le \$200$

$n \le 100$

$= \boxed{ME(100)}$



Stratified problem:

Issue: **Both** costs and precision can depend on how we **allocate** our overall sample size to each stratum

- Strata may be more/less costly to sample
- Measurements within stratum may have different SDs S_h

- The **allocation** fraction for stratum h is

$$a_h = \frac{n_h}{n} \qquad n_h = n a_h$$

- Must have $\sum_{h=1}^{H} a_h = 1$

Stratified example

- $H = 3$ strata with $N_h = 1000$ and $S_h = 1$

Cost: costs per unit in each stratum are $c_1 \overset{\$}{=} 1, c_2 \overset{\$}{=} 2, c_3 \overset{\$}{=} 3$

$$\underline{\text{total cost}} = C(n, a_1, a_2) = \$1\underbrace{a_1 n}_{n_1 = na_1} + \$2\underbrace{a_2 n}_{n_2} + \$3\underbrace{(1 - a_1 - a_2)}_{a_3 = 1 - a_1 - a_2}n$$

Precision: 95% margin of error for estimating the mean

$$ME(n, a_1, a_2) = 1.96 \times \sqrt{\sum_{h=1}^{3} \left(\frac{1000}{3000}\right)^2 \left(1 - \frac{a_h n}{1000}\right) \frac{1}{a_h n}}$$

$$n_h = a_h n$$

Stratified example: determine the n, a_1, a_2 that...

$\left.\begin{array}{l} \min ME(n, a_1, a_2) \\ \{C(n, a_1, a_2) = \$200\} \end{array}\right\}$

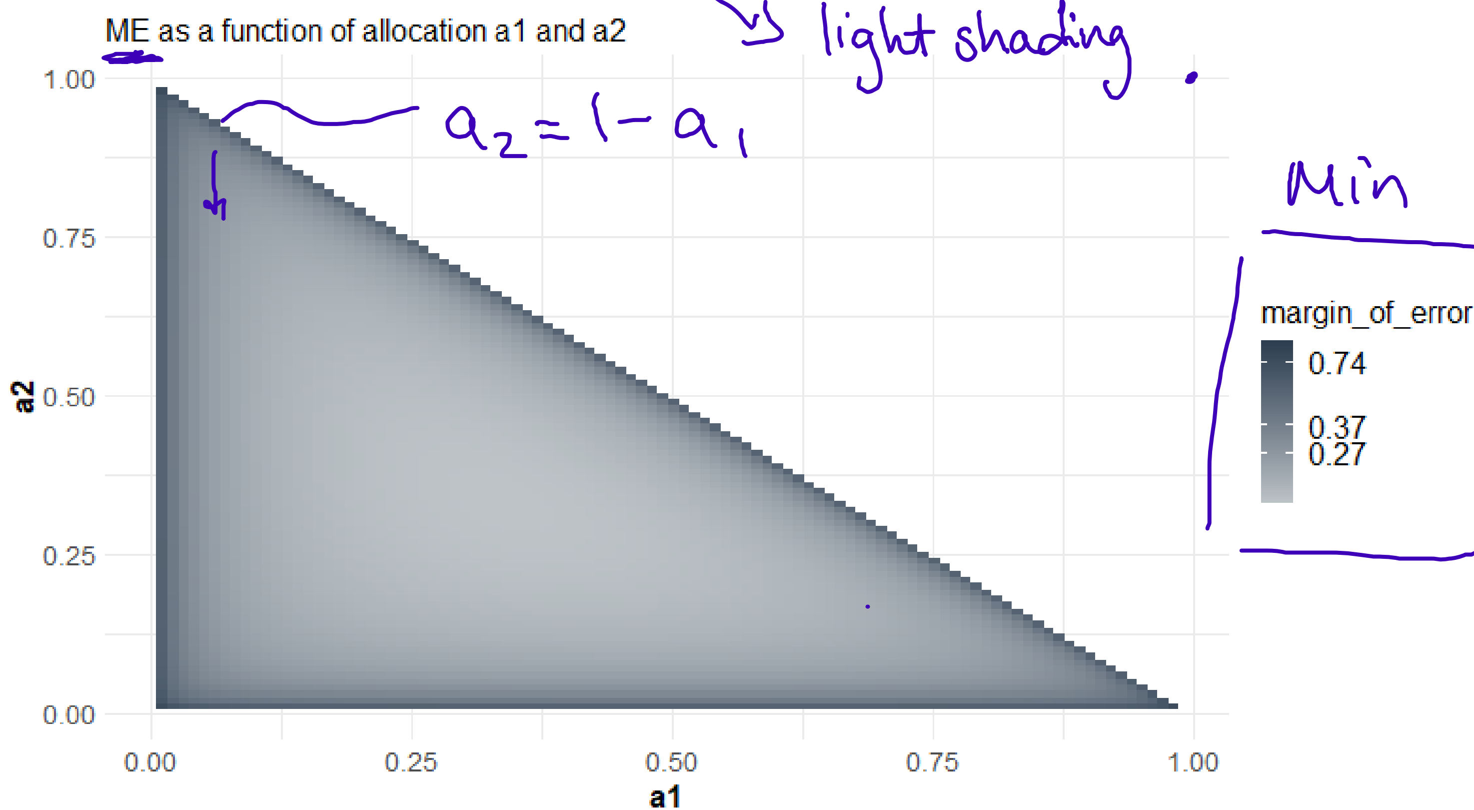
Constraint: Costs equal to \$200

Optimize: minimize margin of error (SE) under this constraint

light shading

$a_2 = 1 - a_1$

Min



Stratified example: determine the n, a_1, a_2 that...

Constraint: Costs equal to \$200

$$\$200 = 1(a_1 n) + 2(a_2 n) + 3(1 - a_1 - a_2)n$$

⋮ algebra (solve for a_2)

$$\boxed{a_2 = \left(3 - \frac{200}{n}\right) - 2a_1}$$

line slope $= -2$

int $= 3 - \frac{200}{n}$

$$0 \le a_1, a_2 \le 1$$

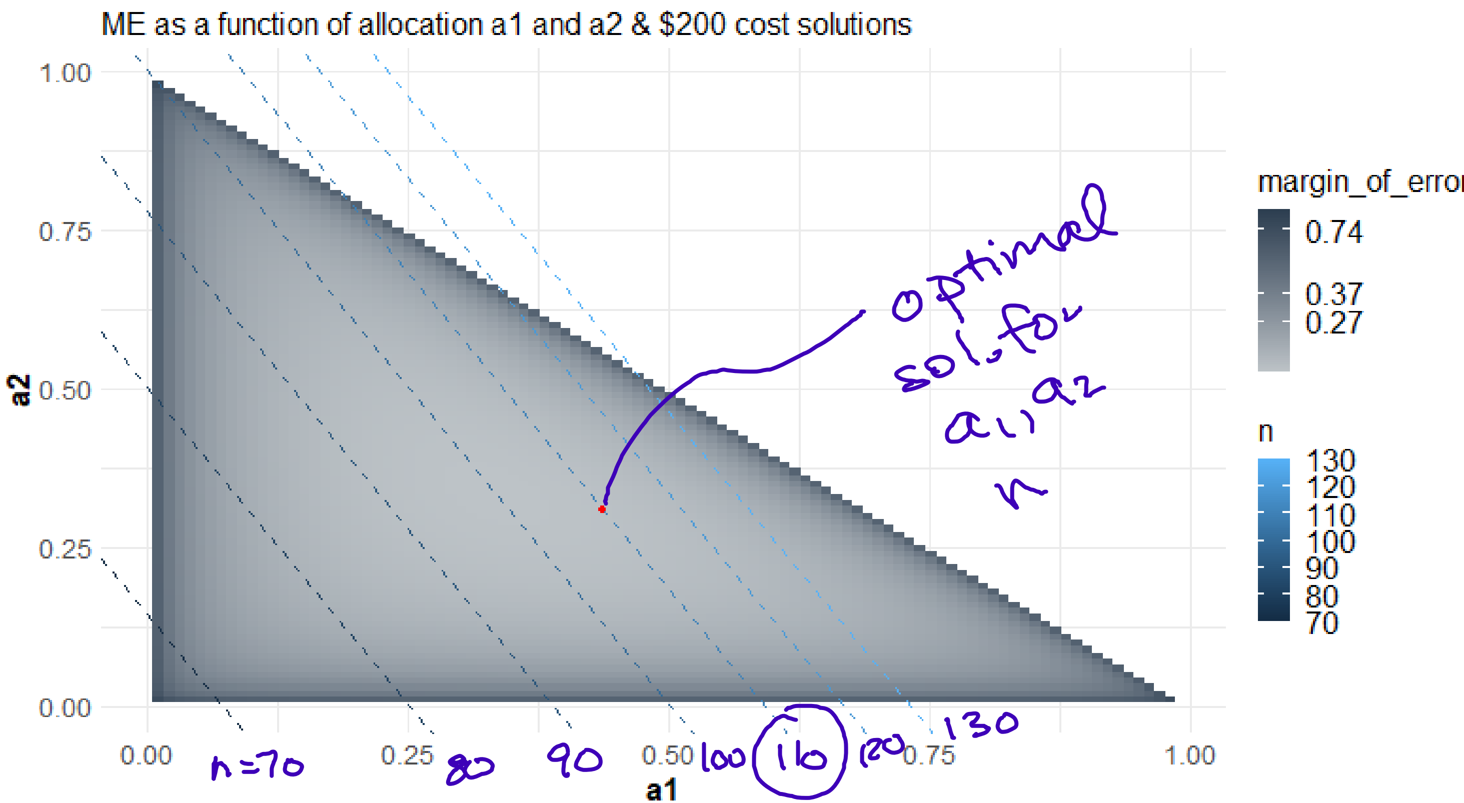
$$0 \le 1 - a_1 - a_2 \le 1$$

$$0 \le a_2 \le 1 - a_1$$

Stratified example: determine the n, a_1, a_2 that...

Constraint: Costs equal to \$200 $\Rightarrow a_2 = \left(3 - \frac{200}{n}\right) - 2a_1$

Optimize: minimize margin of error (SE) under this constraint



Stratified example: determine the n, a_1, a_2 that...

Constraint: Costs equal to $200

Optimize: minimize margin of error (SE) under this constraint

Sol.: LaGrange Mult. Method

gives sol.: $a_h = \dfrac{N_h S_h / \sqrt{c_h}}{\sum_h N_h S_h / \sqrt{c_h}}$

$a_1 = \dfrac{1000(1)/\sqrt{1}}{1000(1)/\sqrt{1} + 1000(1)/\sqrt{2} + 1000(1)/\sqrt{3}} \approx \boxed{.4377}$

$a_2 = \boxed{.3095}$ $a_3 = \boxed{.2528}$

$\$200 = \$1(.4377)n + \$2(.3095)n + \$3(.2528)n$

solve for n: $n = 110.19 \Rightarrow \boxed{n = 110}$ to not go over budget

Design opt. for budget of \$200:

$$n_1 = 110(.4377) = 48.18 \approx 48$$
$$n_2 = 110(.3095) = 34.05 \approx 34$$
$$n_3 = 110(.2528) = 27.80 \approx 28$$
$$\overline{110}$$

$\$1(48) + \$2(34) + \$3(28) = \$200$ ✓

$\boxed{ME = .188}$ → no other alloc. will give a smaller ME for a cost of \$200.

Compare to prop. alloc: $a_1 = a_2 = a_3 = \frac{1000}{3000} = \frac{1}{3}$

cost = \$200 $\quad n = \dfrac{\$200}{\$1(\frac{1}{3}) + \$2(\frac{1}{3}) + \$3(\frac{1}{3})} = \boxed{100}$

for cost of \$200, sample 100 units & str. ME is equal to $\boxed{.193}$ → larger than opt. alloc. solution