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...

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

Optimize: minimize cost under this constraint

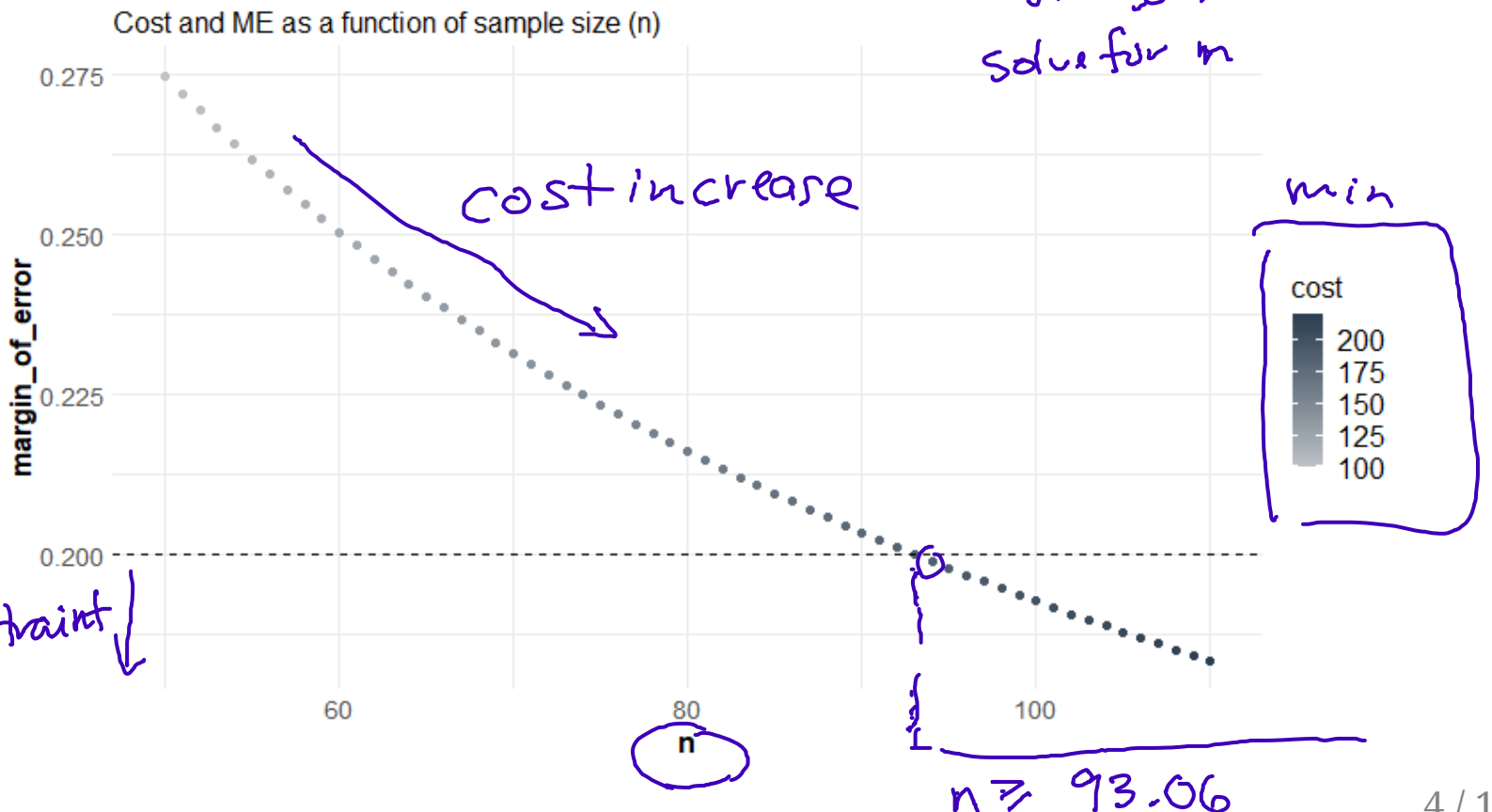
$\$2 \cdot n$

$\min C(n) = C(94)$

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

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

solve for n



SRS example: determine the n that...

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

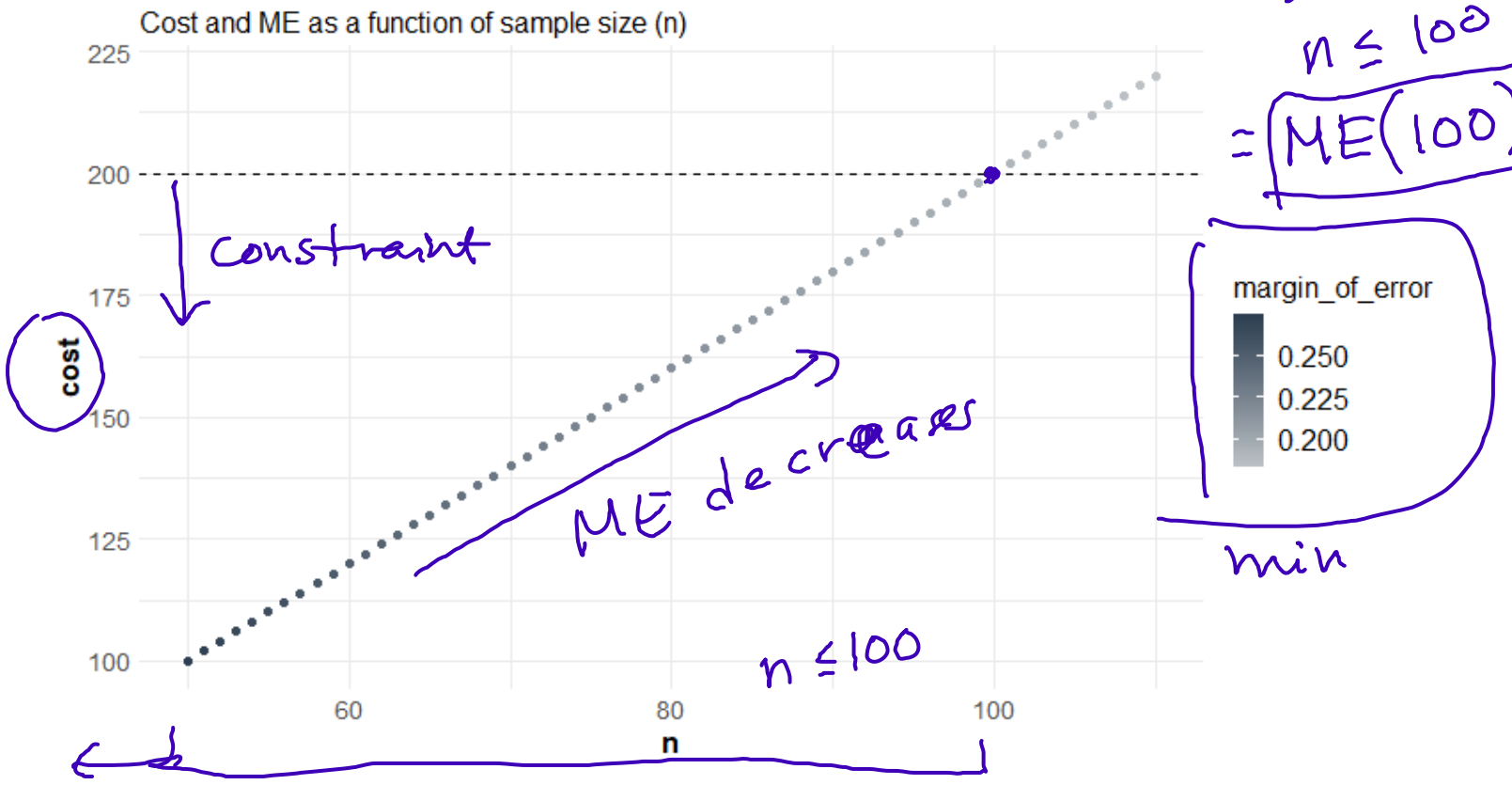
Optimize: minimize margin of error (SE) under this constraint

$\min ME(n)$ $\{C(n) \leq \$200\}$

$\$2 \cdot n \leq \$200$

$n \leq 100$

$= ME(100)$



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 $c_1 = 1, c_2 = 2, c_3 = 3$

$$\text{total cost} = C(n, a_1, a_2) = \$1 a_1 n + \$2 a_2 n + \$3(1 - a_1 - a_2)n$$

$n_1 = n a_1$ n_2 $a_3 = 1 - a_1 - a_2$

Precision: 95% margin of error for estimating the mean

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

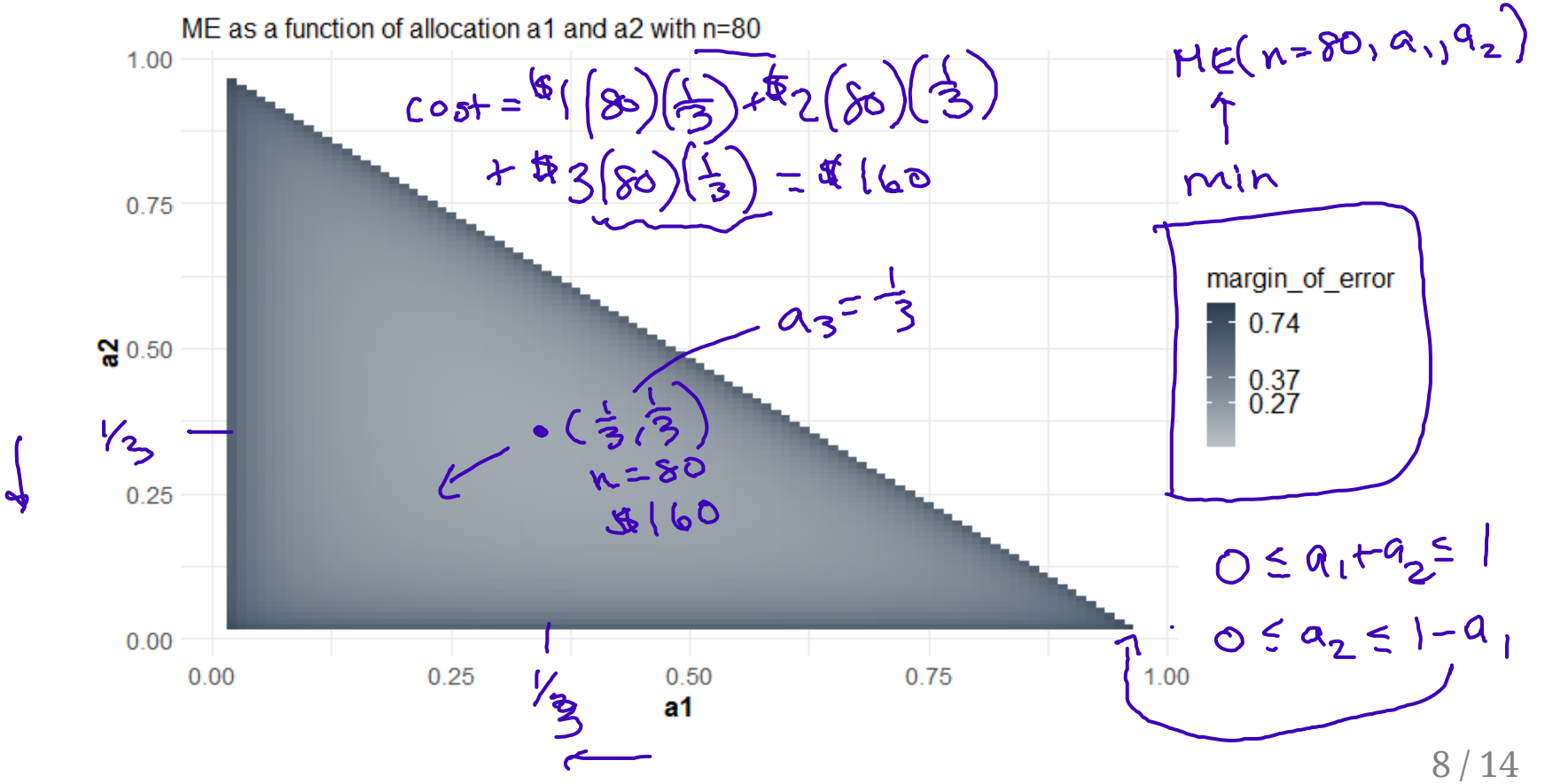
$n_h = a_h n$

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

* **Constraint:** Costs equal to \$200 *

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

$\min ME(n, a_1, a_2)$ subject to $C(n, a_1, a_2) = 200$



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

Constraint: Costs equal to \$200

$$\$200 = \$1(na_1) + \$2(na_2) + \$3(n)(1-a_1-a_2)$$

solve for a_2 ... algebra

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

intercept; slope -2

e.g. $n = 80$

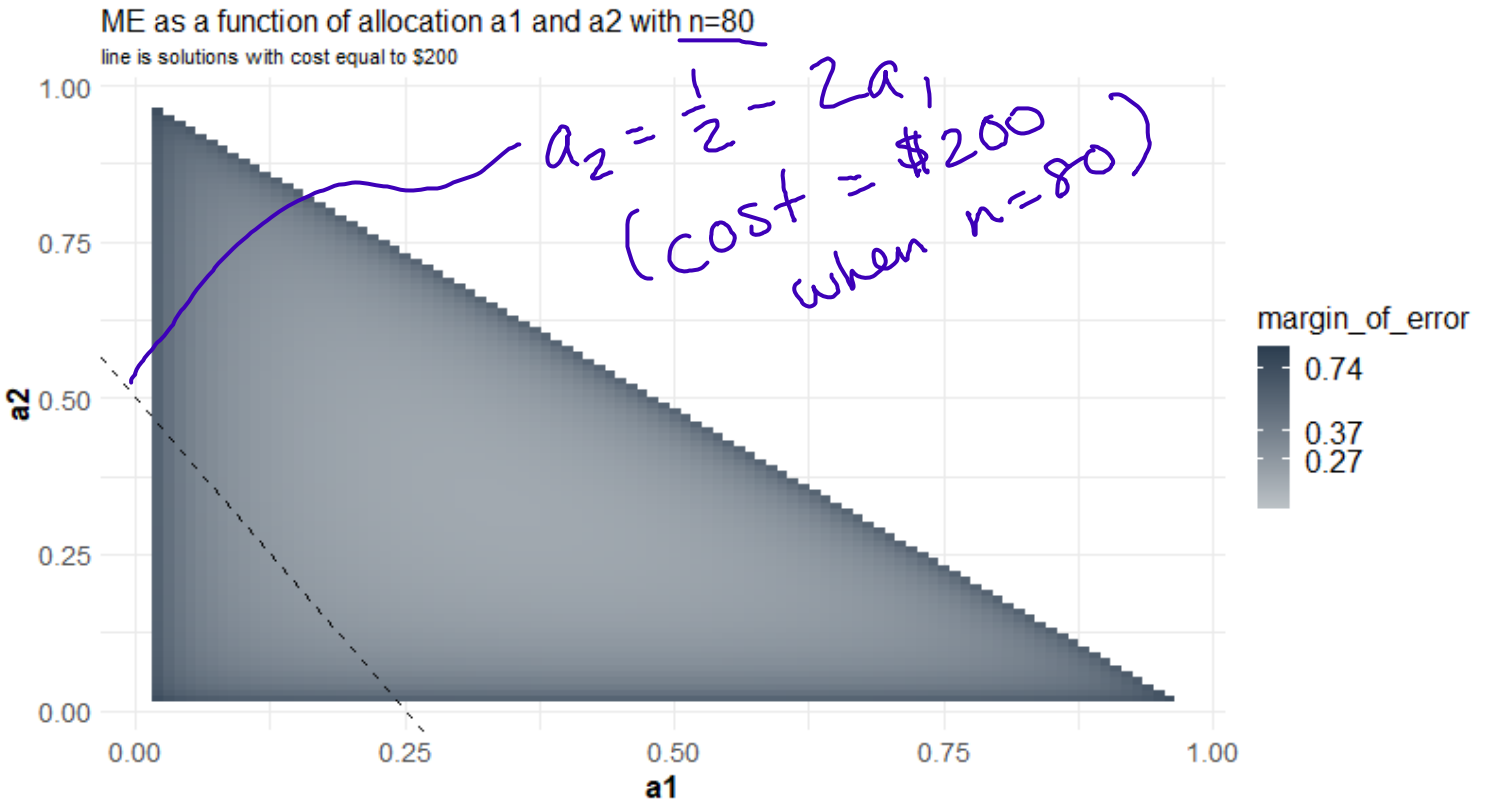
constraint is

$$a_2 = \left(3 - \frac{200}{80}\right) - 2a_1 = .5 - 2a_1$$

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

Constraint: Costs equal to $200

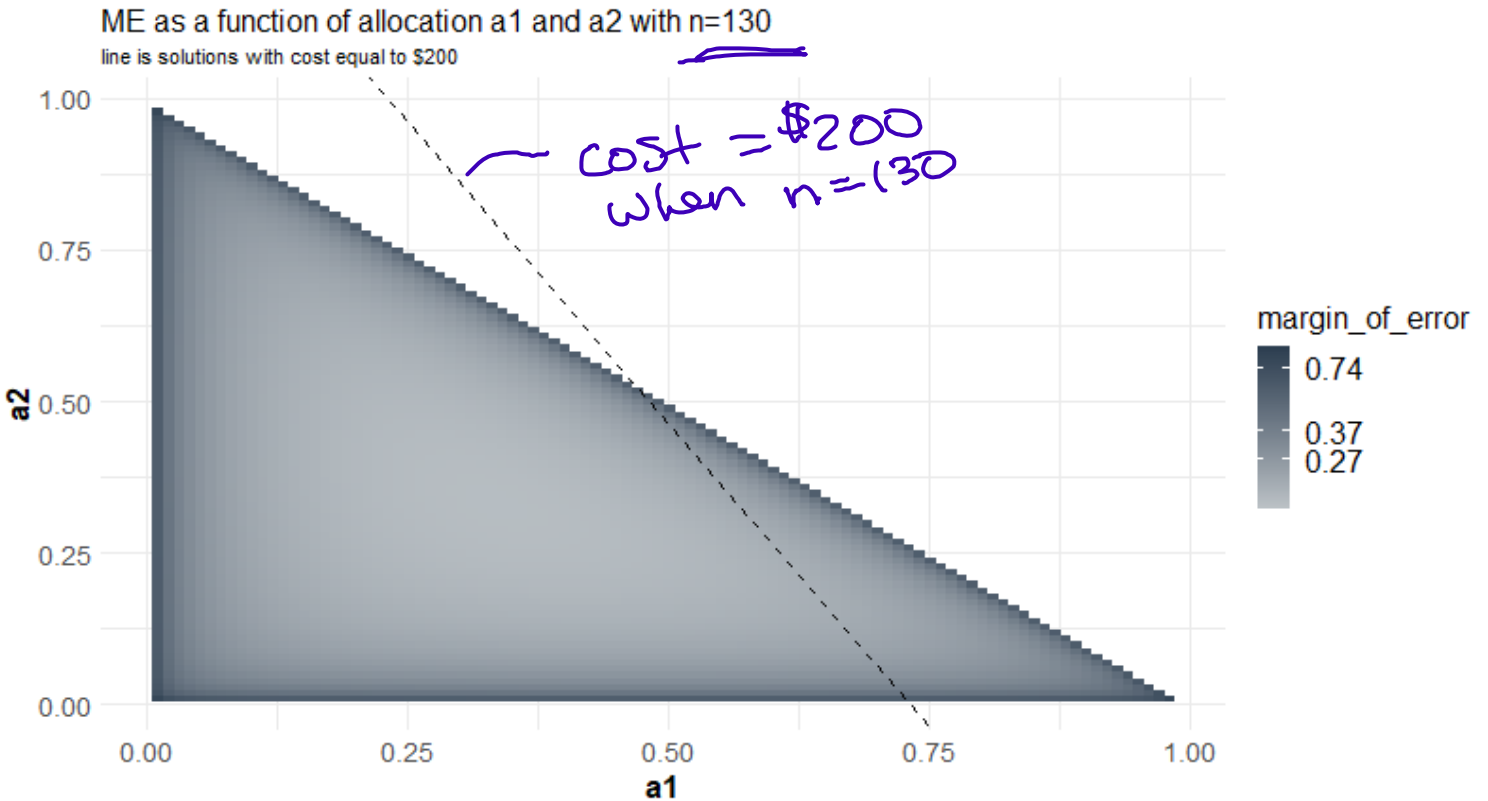
Optimize: minimize margin of error (SE) under this constraint



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

- **Constraint:** Costs equal to \$200

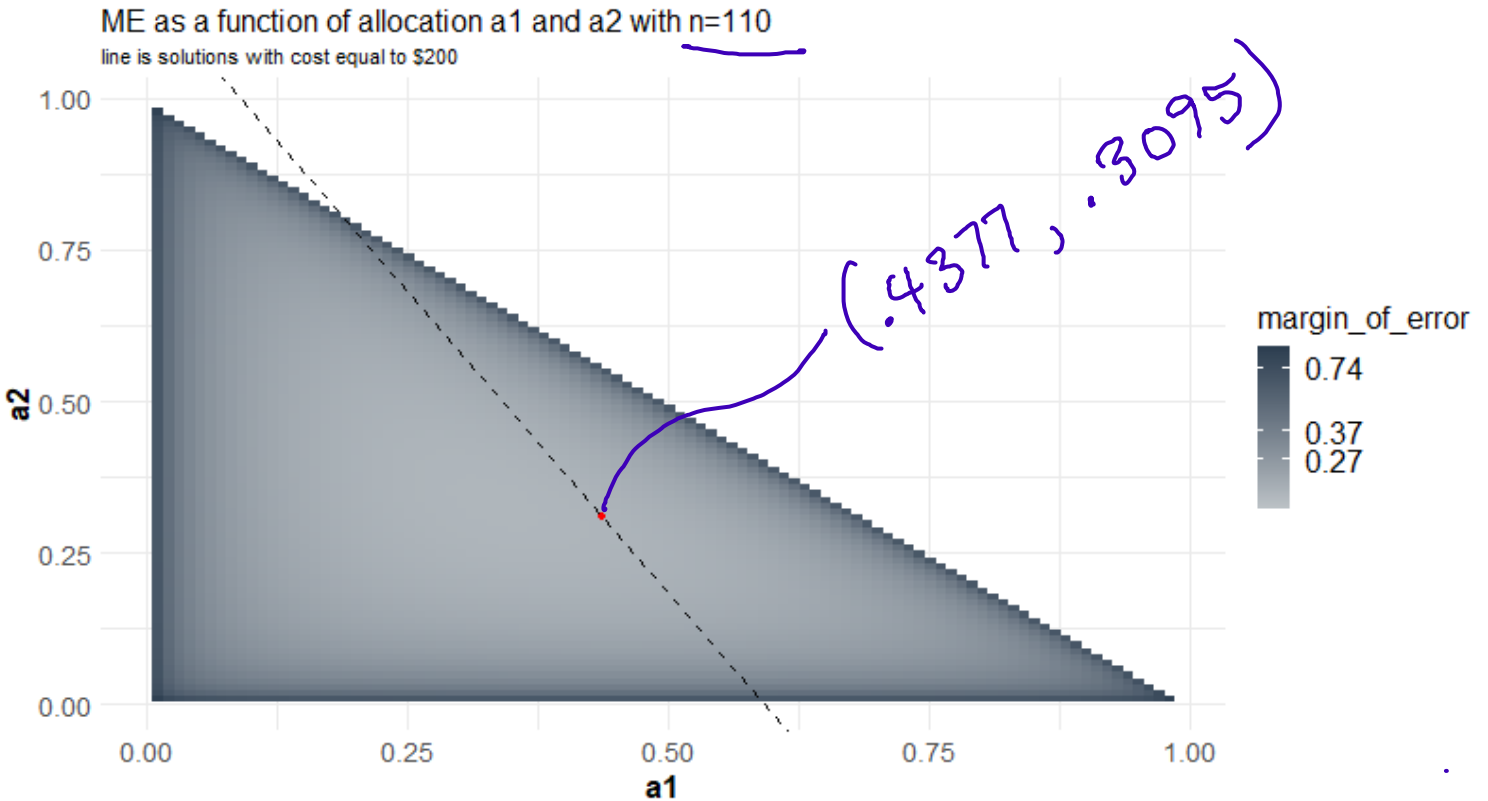
Optimize: minimize margin of error (SE) under this constraint



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

- **Constraint:** Costs equal to \$200

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

LaGrange Multiplier Method

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

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

$a_2 = .3095$ $\quad a_3 = .2528$

$n = ??$

get n

Constraint : cost $\leq \$200$

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

$$n = \frac{\$200}{\$1(.4377) + \$2(.3095) + \$3(.2528)} \approx 110.186$$

$n = 110$ (don't go over cost)

try $q_1 = q_2 = q_3 = \frac{1}{3}$ (prop. alloc)
cost = \$200
n = 100
ME = .193

$n_1 = 110(.4377) \approx 48.18 \approx 48$

$n_2 = 110(.3095) \approx 34.05 \approx 34$

$n_3 = 110(.2528) \approx 27.8 \approx 28$

110

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

$ME(n=110, q_1=.4377, q_2=.3095) \simeq .188$ → smallest ME possible when spending \$200