# Jason Abaluck
optimal Representation
Ephraim Sutherland

# Contents

## Setup

1. Suppose a physician can only see see ATE and some measure of representativeness. They have prior $\bar{\beta}$ and $\beta_{ATE} = (1/N) \sum \beta_i$.

2. need model for betas related to each other based on $x$'s. WLOG, suppose

$$\beta(x_i) = x_i \gamma$$

   Where $x_i$ is a vector of characteristics and $\gamma$ is a vector of coefficients.
   If you know $\gamma$, then you know $\beta$ for any given patient.

3. However, you don't observe $\gamma$, you instead observe: $\beta_{ATE} = \bar{x}\gamma$ where $\bar{x} = \left(\frac{1}{N}\right) \sum x_i$

4. We know $\beta_i$ for patients with characteristics $\bar{x}$ (it is $\beta_{ATE}$).

5. For other patients, need to solve

$$\beta_{i,post} = \mathrm{E}(x_i \gamma | \bar{x}\gamma = \beta_{ATE})$$

6. to solve

    (a)

$$\beta_{i,post} = \mathrm{E}(x_i \gamma | \bar{x}\gamma = \beta_{ATE})$$

$$= \mathrm{E}((x_i - c_i\bar{x})\gamma | \bar{x}\gamma = \beta_{ATE}) + c_i\mathrm{E}(\bar{x}\gamma | \bar{x}\gamma = \beta_{ATE})$$

$$= \mathrm{E}((x_i - c_i\bar{x})\gamma | \bar{x}\gamma = \beta_{ATE}) + c_i\beta_{ATE}$$

   For any constant $c_i$.
   Choose $c_i$ so that

$$\mathrm{Cov}((x_i - c_i\bar{x})\gamma, \bar{x}\gamma) = 0$$

   maybe assume normality so that this guarantees independence. Then,

$$\mathrm{E}((x_i - c_i\bar{x})\gamma | \bar{x}\gamma = \beta_{ATE}) = (x_i - c_i\bar{x})\mathrm{E}(\gamma)$$

So then

$$(x_i - c_i\bar{x})\mathrm{E}(\gamma) + c_i\beta_{ATE} = x_i\mathrm{E}(\gamma) + c_i(\beta_{ATE} - \bar{x}\mathrm{E}(\gamma))$$

($c_i$ depends on $x_i$)

In other words, your belief is your prior, adjusted based on the difference between the observed ATE and your prior about the ATE. The key question is how much adjustment you do which depends on "$c_i$". We choose $c_i$ to solve:

$$\mathrm{Cov}((x_i - c_i\bar{x})\gamma, \bar{x}\gamma) = 0$$

$$\Longleftrightarrow \mathrm{Cov}(x_i\gamma, \bar{x}\gamma) - c_i\mathrm{Cov}(\bar{x}\gamma, \bar{x}\gamma) = 0$$

$$\Longleftrightarrow \mathrm{Cov}(x_i\gamma, \bar{x}\gamma) = c_i\mathrm{Var}(\bar{x}\gamma)$$

$$\Longleftrightarrow c_i = \frac{\mathrm{Cov}(\beta_i, \beta_{ATE})}{\mathrm{Var}(\beta_{ATE})}$$

The random variable in this context is $\gamma$ (the coefficients on the $x$'s) in this case $\mathrm{Var}(\beta_{ATE})$ is a measure of how uncertain one was about what $\beta_{ATE}$ would be before doing the trial.

$c_i$ is the equation for a regression of $\beta_i$ on $\beta_{ATE}$. In other words, we take a bunch of patients with characteristics $x_i$ and we keep redrawing the gammas from our prior distribution the we ask how correlated $\beta_i$ and $\beta_{ATE}$ are. If they are more correlated (as they would be for patients where the $x_i$ are closer to $\bar{x}$ we update more.

To compute $c_i$, we just need to know $x_i$ $\bar{x}$, and the distribution of $\gamma$.

Suppose we want to design the trial to minimize:

$$\min \mathrm{E}[(\beta_i - \beta_{i,post})^2]$$

## Simple Cases

1. There is just one $x$ and it is binary (old v young). Can it be solved analytically?

2. Can you solve a 2-dimensional case?

First observe that in our current setup, $c$ does not depend on the $\gamma$'s
We have that

$$c = \frac{\mathrm{Cov}(x\gamma, \bar{x}\gamma)}{\mathrm{Var}(\bar{x}\gamma)}$$

so for individual i, $c$ reduces to

$$c = \frac{1^2\mathrm{Var}(\gamma_0) + x\bar{x}\mathrm{Var}(\gamma_1)}{1^2\mathrm{Var}(\gamma_0) + \bar{x}^2\mathrm{Var}(\gamma_1)} \quad \text{because } \mathrm{Var}(\gamma_0) = \mathrm{Var}(\gamma_1) = \mathrm{Var}(\gamma) = 1$$

$$= \frac{(1 + x\bar{x})\mathrm{Var}(\gamma)}{(1 + \bar{x}^2)\mathrm{Var}(\gamma)} = \frac{\vec{x}'\vec{\bar{x}}}{\vec{\bar{x}}'\vec{\bar{x}}}$$

# 1 Appendix

## Derivations

## 1. symmetric case

Recall we want to minimize

$$\min \mathrm{E}_x \left( \mathrm{E}_{\gamma_{0,1}}[(\beta_i - \beta_{i,post})^2] \right)$$

One way we can rewrite these equations is as the effect of women vs men. let the squared error (SE) be

$$\sqrt{SE} = \beta_i - \beta_{i,post} = \underbrace{[(1-x)\gamma_0 + x\gamma_1]}_{\beta_i} - \underbrace{[((\vec{x} - c\vec{\bar{x}})\mathrm{E}(\gamma) + c\beta_{ATE})]}_{\beta_{i,post}}$$

$$= [(1-x)\gamma_0 + x\gamma_1] - [((1-x) - c(1-\bar{x}))\mathrm{E}(\gamma_0) + (x - c\bar{x})\mathrm{E}(\gamma_1) + c((1-\bar{x})\gamma_0 + \bar{x}\gamma_1)]$$

$\beta_i$ $\beta_{i,post}$ and as a result, $SE$ are all a function of $x$. We can then describe $\beta_i^{men} = \beta_i(x = 1)$ and similarly for other terms to get
So

$$\beta_i^{men} = \gamma_1$$

$$\beta_i^{post,men} = -c_{men}(1-\bar{x})\bar{\gamma}_0 + (1 - c_{men}\bar{x})\bar{\gamma}_1 + c_{men}[(1-\bar{x})\gamma_0 + \bar{x}\gamma_1]$$

$$\beta_i^{women} = \gamma_0$$

$$\beta_i^{post,women} = [1 - c_{wom}(1-\bar{x})]\mathrm{E}(\gamma_0) - c_{wom}\bar{x}\mathrm{E}(\gamma_1) + c_{wom}(1-\bar{x})\gamma_0 + c_{wom}\bar{x}\gamma_1$$

this can be broken down into

$$\sqrt{SE} = (\beta_i^{men} - \beta_i^{post,men}) + (\beta_i^{wom} - \beta_i^{post,wom})$$

let

$$W^{men} = (\beta_i^{men} - \beta_i^{post,men})$$

$$= (1 - \bar{x}c_{men})(\gamma_1 - \bar{\gamma}_1) - c_{men}(1 - \bar{x})(\gamma_0 - \bar{\gamma}_0)$$

$$= (c_{wom}(1 - \bar{x}))(\gamma_1 - \bar{\gamma}_1) - c_{men}(1 - \bar{x})(\gamma_0 - \bar{\gamma}_0)$$

$$W^{wom} = (\beta_i^{wom} - \beta_i^{post,wom})$$

$$= [1 - c_{wom}(1 - \bar{x})](\gamma_0 - \bar{\gamma}_0) - c_{wom}\bar{x}(\gamma_1 - \bar{\gamma}_1)$$

$$= [c_{men}\bar{x}](\gamma_0 - \bar{\gamma}_0) - c_{wom}\bar{x}(\gamma_1 - \bar{\gamma}_1)$$

and recall that

$$c_i = \frac{(1 - x)(1 - \bar{x})\text{Var}(\gamma_0) + x\bar{x}\text{Var}(\gamma_1)}{(1 - \bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}$$

so if you are a man, then

$$c_{men} = \frac{\bar{x}\text{Var}(\gamma_1)}{(1 - \bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}$$

and likewise if you are a woman, then

$$c_{woman} = \frac{(1 - \bar{x})\text{Var}(\gamma_0)}{(1 - \bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}$$

And, allowing for arbitrary variances, we can say that

$$W^2_{men} = \left( \frac{(1-\bar{x})^4}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \right) \text{Var}(\gamma_0)^2 \text{Var}(\gamma_1)$$

$$- \frac{2(1-\bar{x})^2[\bar{x}(1-\bar{x})]}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)\text{Var}(\gamma_0)\text{Cov}(\gamma_0, \gamma_1)$$

$$+ \frac{\bar{x}^2(1-\bar{x})^2}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)^2 \text{Var}(\gamma_0)$$

$$= \text{Var}(\gamma_0)\text{Var}(\gamma_1) \left( \frac{(1-\bar{x})^4}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0) \right.$$

$$- \frac{2(1-\bar{x})^2[\bar{x}(1-\bar{x})]}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Cov}(\gamma_0, \gamma_1)$$

$$\left. + \frac{\bar{x}^2(1-\bar{x})^2}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1) \right)$$

$$W^2_{women} = \left( \frac{(\bar{x})^4}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \right) \text{Var}(\gamma_1)^2 \text{Var}(\gamma_0)$$

$$- \frac{2\bar{x}^2[\bar{x}(1-\bar{x})]}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)\text{Var}(\gamma_0)\text{Cov}(\gamma_0, \gamma_1)$$

$$+ \frac{\bar{x}^2(1-\bar{x})^2}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0)^2 \text{Var}(\gamma_1)$$

$$= \text{Var}(\gamma_0)\text{Var}(\gamma_1) \left( \frac{(\bar{x})^4}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1) \right.$$

$$- \frac{2\bar{x}^2[\bar{x}(1-\bar{x})]}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Cov}(\gamma_0, \gamma_1)$$

$$\left. + \frac{\bar{x}^2(1-\bar{x})^2}{((1-\bar{x})^2 \text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0) \right)$$

So letting $\text{Cov}(\gamma_0, \gamma_1) = 0$ we get

$$W^2_{men} = \text{Var}(\gamma_0)\text{Var}(\gamma_1)\left(\frac{(1-\bar{x})^2}{(1-\bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}\right)$$

$$W^2_{women} = \text{Var}(\gamma_0)\text{Var}(\gamma_1)\left(\frac{(\bar{x})^2}{(1-\bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}\right)$$

Now observe that if $\alpha = \text{Var}(\gamma_1)$ and $\beta = \text{Var}(\gamma_0)$ then

$$\alpha W^2_{women} + \beta W^2_{men} = \text{Var}(\gamma_0)\text{Var}(\gamma_1)\left[\frac{\text{Var}(\gamma_1)\bar{x}^2 + \text{Var}(\gamma_0)(1-\bar{x})^2}{(1-\bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}\right]$$

$$= \text{Var}(\gamma_0)\text{Var}(\gamma_1)$$

So we can write

$$\alpha W^2_{women} = \text{Var}(\gamma_0)\text{Var}(\gamma_1) - \beta W^2_{men}$$

$$W^2_{women} = \frac{\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \beta W^2_{men}}{\alpha}$$

$$W^2_{women} = \frac{\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Var}(\gamma_0)W^2_{men}}{\text{Var}(\gamma_1)}$$

And thus conclude

$$MSE = (1-p)W^2_{women} + pW^2_{men}$$

$$MSE = pW^2_{men} + (1-p)\left(\frac{\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Var}(\gamma_0)W^2_{men}}{\text{Var}(\gamma_1)}\right)$$

$$MSE = pW^2_{men} + (1-p)\left(\frac{\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Var}(\gamma_0)W^2_{men}}{\text{Var}(\gamma_1)}\right)$$

$$= \frac{\left(p[\text{Var}(\gamma_0) + \text{Var}(\gamma_1)] - \text{Var}(\gamma_0)\right)W^2_{men} + (1-p)\text{Var}(\gamma_0)\text{Var}(\gamma_1)}{\text{Var}(\gamma_1)}$$

Thus when $p > \frac{\text{Var}(\gamma_0)}{\text{Var}(\gamma_0)+\text{Var}(\gamma_1)}$ we can clearly see that MSE is minimized when $W^2_{men}$ is minimized (when $\bar{x} = 1$. And inversely when $p < \frac{\text{Var}(\gamma_0)}{\text{Var}(\gamma_0)+\text{Var}(\gamma_1)}$ MSE is minimized when $W^2_{men}$

is maximaximized (when $\bar{x} = 0$. In other words, when the proportion of men is $p > \frac{\text{Var}(\gamma_0)}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)}$ it is optimal to have only men ($\bar{x}$) in the trial, and vice versa. And when there are eqal number of men and women in the population, $MSE$ does not depend on $W^2_{men}$ and thus has equal error of $\frac{\text{Var}(\gamma_0)}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)}$ for any $\bar{x}$.

We can also solve this using the first order conditions.

$$MSE = (1-p)W^2_{women} + pW^2_{men} = \text{Var}(\gamma_0)\text{Var}(\gamma_1)\Big[\frac{(1-p)\bar{x}^2 + p(1-\bar{x})^2}{(1-\bar{x})^2\text{Var}(\gamma_0) + \bar{x}^2\text{Var}(\gamma_1)}\Big]$$

Which gives a derivative of

$$\frac{MSE}{d\bar{x}} = -\frac{2\text{Var}(\gamma_0)\text{Var}(\gamma_1)(1-\bar{x})\bar{x}\big[\text{Var}(\gamma_0)(p-1) + \text{Var}(\gamma_1)p\big]}{(\text{Var}(\gamma_0)(\bar{x}-1)^2 + \text{Var}(\gamma_1)\bar{x}^2)^2}$$

Here we can see that we still have roots $\bar{x} = 0$ and $\bar{x} = 1$ and all that changes is we get a variances-weighted edge-case whenever $(\text{Var}(\gamma_0) + \text{Var}(\gamma_1))p - \text{Var}(\gamma_0) = 0$

And to get which is the minizing solution we can observe the SOC

$$\frac{MSE}{d\bar{x}^2} = -\frac{2\text{Var}(\gamma_0)\text{Var}(\gamma_1)[\text{Var}(\gamma_0)(p-1) + \text{Var}(\gamma_1)p][\text{Var}(\gamma_0)(2x+1)(x-1)^2 + \text{Var}(\gamma_1)x^2(2x-3)]}{[\text{Var}(\gamma_0)(x-1)^2 + \text{Var}(\gamma_1)x^2]^3}$$

From the second order condition, we can look at the two roots. We can see that the numerator reduces to two cases. When $p > \frac{\text{Var}(\gamma_0)}{(\text{Var}(\gamma_0) + \text{Var}(\gamma_1))}$, then we can observe that the numerator is positive for the root $\bar{x} = 1$. Conversely, the numerator is positive when $\bar{x} = 0$ We can also see this outlined in the simulations below.

## 2. Just Men

$$\beta_i = \gamma_0 + x\gamma_1$$

$$\beta_i^{post} = \text{E}(\gamma_0) + \text{E}(\gamma_1) + c_i(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

$$\beta_i^{women} = \gamma_0$$

$$\beta_i^{men} = \gamma_0 + \gamma_1$$

$$\beta_i^{post,men} = \text{E}(\gamma_0) + \text{E}(\gamma_1) + c_{men}(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

$$\beta_i^{post,women} = \text{E}(\gamma_0) + c_{women}(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

And we have

$$C_i = \frac{\mathrm{Var}(\gamma_0) + x\bar{x}\mathrm{Var}(\gamma_1)}{\mathrm{Var}(\gamma_0) + \bar{x}^2\mathrm{Var}\gamma_1}$$

$$C_{women} = \frac{\mathrm{Var}(\gamma_0)}{\mathrm{Var}(\gamma_0) + \bar{x}^2\mathrm{Var}(\gamma_1)}$$

$$C_{men} = \frac{\mathrm{Var}(\gamma_0) + \bar{x}\mathrm{Var}(\gamma_1)}{\mathrm{Var}(\gamma_0) + \bar{x}^2\mathrm{Var}(\gamma_1)}$$

$$W_{women} = \beta_i^{women} - \beta_i^{post,women}$$

$$= (\gamma_0 - \mathrm{E}(\gamma_0)) - C_{men}\big[(\gamma_0 - \mathrm{E}(\gamma_0)) + \bar{x}(\gamma_1 - \mathrm{E}(\gamma_1))\big]$$

$$= (1 - c_{women})(\gamma_0 - \mathrm{E}(\gamma_0)) - C_{women}\bar{x}(\gamma_1 - \mathrm{E}(\gamma_1))$$

$$W_{men} = \beta_i^{men} - \beta_i^{post,men}$$

$$= (\gamma_0 - \mathrm{E}(\gamma_0)) + (\gamma_1 - \mathrm{E}(\gamma_1) - C_{men}\big[(\gamma_0 - \mathrm{E}(\gamma_0)) + \bar{x}(\gamma_1 - \mathrm{E}(\gamma_1))\big]$$

$$= (1 - c_{men})(\gamma_0 - \mathrm{E}(\gamma_0)) + (1 - C_{men}\bar{x})(\gamma_1 - \mathrm{E}(\gamma_1))$$

Squaring

$$W_{women}^2 = (1 - c_{wom})^2\mathrm{Var}(\gamma_0) - 2(1 - c_{wom})c_{wom}\bar{x}\mathrm{Cov}(\gamma_0, \gamma_1) + c_{wom}^2\bar{x}\mathrm{Var}(\gamma_1)$$

$$W_{men}^2 = (1 - c_{men})^2\mathrm{Var}(\gamma_0) + 2(1 - c_{men})(1 - c_{men}\bar{x})\mathrm{Cov}(\gamma_0, \gamma_1) + (1 - c_{men}\bar{x})^2\mathrm{Var}(\gamma_1)$$

And assuming $\mathrm{Cov}(\gamma_0, \gamma_1) = 0$ we get

$$W_{women}^2 = (1 - c_{wom})^2 \text{Var}(\gamma_0) + c_{wom}^2 \bar{x}^2 \text{Var}(\gamma_1)$$

$$= \frac{\bar{x}^4 \text{Var}(\gamma_1)^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0) + \frac{\bar{x}^2 \text{Var}(\gamma_0)^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)$$

$$= \bar{x}^2 \text{Var}(\gamma_1) \text{Var}(\gamma_0) \frac{\bar{x}^2 \text{Var}(\gamma_1) + \text{Var}(\gamma_0)}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2}$$

$$= \frac{\bar{x}^2 \text{Var}(\gamma_1) \text{Var}(\gamma_0)}{\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1)}$$

$$W_{men}^2 = (1 - c_{men})^2 \text{Var}(\gamma_0) + (1 - c_{men}\bar{x})^2 \text{Var}(\gamma_1)$$

$$= \frac{\text{Var}(\gamma_1)^2 [\bar{x}^2 - \bar{x}]^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0) + \frac{\text{Var}(\gamma_0)^2 [1 - \bar{x}]^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)$$

$$= \frac{\text{Var}(\gamma_1)^2 \bar{x}^2 [\bar{x} - 1]^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_0) + \frac{\text{Var}(\gamma_0)^2 [\bar{x} - 1]^2}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2} \text{Var}(\gamma_1)$$

$$= \text{Var}(\gamma_0) \text{Var}(\gamma_1) [\bar{x} - 1]^2 \frac{\text{Var}(\gamma_1)\bar{x}^2 + \text{Var}(\gamma_0)}{(\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1))^2}$$

$$= \frac{\text{Var}(\gamma_0) \text{Var}(\gamma_1) [\bar{x} - 1]^2}{\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1)}$$

So we get MSE of

$$MSE = \text{Var}(\gamma_0) \text{Var}(\gamma_1) \frac{(1 - p)\bar{x}^2 + p[\bar{x} - 1]^2}{\text{Var}(\gamma_0) + \bar{x}^2 \text{Var}(\gamma_1)}$$

Taking first order conditions we get

$$\frac{d}{dx} \frac{\text{Var}(\gamma_0) \text{Var}(\gamma_1) [(1 - p)x^2 + p(x - 1)^2]}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)x^2}$$

$$= 2\text{Var}(\gamma_0) \text{Var}(\gamma_1) \frac{\text{Var}(\gamma_0)(x - p) + \text{Var}(\gamma_1)p(x - 1)x}{(\text{Var}(\gamma_0) + \text{Var}(\gamma_1)x^2)^2} = 0$$

$$\iff \text{Var}(\gamma_0)(x - p) + \text{Var}(\gamma_1)p(x - 1)x = 0$$

$$\iff x^2 \text{Var}(\gamma_1)p + x(\text{Var}(\gamma_0) - \text{Var}(\gamma_1)p) - \text{Var}(\gamma_0)p = 0$$

For this we can use the quadratic formula with

1. $a = \text{Var}(\gamma_1)p$

2. $b = \text{Var}(\gamma_0) - \text{Var}(\gamma_1)p$

3. $c = -\text{Var}(\gamma_0)p$

giving us

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-(\text{Var}(\gamma_0) - \text{Var}(\gamma_1)p) \pm \sqrt{(\text{Var}(\gamma_0) - \text{Var}(\gamma_1)p)^2 + 4\text{Var}(\gamma_0)\text{Var}(\gamma_1)p^2}}{2\text{Var}(\gamma_1)p}$$

$$= \frac{\text{Var}(\gamma_1)p - \text{Var}(\gamma_0) \pm \sqrt{\text{Var}(\gamma_0)^2 - 2\text{Var}(\gamma_0)\text{Var}(\gamma_1)p + \text{Var}(\gamma_1)^2p^2 + 4\text{Var}(\gamma_0)\text{Var}(\gamma_1)p^2}}{2\text{Var}(\gamma_1)p}$$

$$= \frac{\text{Var}(\gamma_1)p - \text{Var}(\gamma_0) \pm \sqrt{\text{Var}(\gamma_0)^2 + 2\text{Var}(\gamma_0)\text{Var}(\gamma_1)p(2p-1) + \text{Var}(\gamma_1)^2p^2}}{2\text{Var}(\gamma_1)p}$$

We can see the errors and optimal solutions 1 and 1

## 3. Generalization of case 1 and 2

In case 1, we solved

$$\beta_i = (1-x)\gamma_0 + x\gamma_1 = \gamma_0 + x(\gamma_1 - \gamma_0)$$

Thus in effect, the constant and coefficient are correlated through $\gamma_0$. We can thus nest both cases by allowing for arbitrary correlation in case 1.

We will do this below

First for $C$

We have that

$$c = \frac{\text{Cov}(x\gamma, \bar{x}\gamma)}{\text{Var}(\bar{x}\gamma)}$$

so for individual i, $c$ reduces to

$$c = \frac{\text{Var}(\gamma_0) + (x + \bar{x})\text{Cov}(\gamma_0, \gamma_1) + x\bar{x}\text{Var}(\gamma_1)}{\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)}$$

$$C_{women} = \frac{\text{Var}(\gamma_0) + \bar{x}\text{Cov}(\gamma_0, \gamma_1)}{\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)}$$

$$C_{men} = \frac{\text{Var}(\gamma_0) + (1 + \bar{x})\text{Cov}(\gamma_0, \gamma_1) + \bar{x}\text{Var}(\gamma_1)}{\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)}$$

We still have

$$\beta_i = \gamma_0 + x\gamma_1$$

$$\beta_i^{post} = \text{E}(\gamma_0) + \text{E}(\gamma_1) + c_i(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

$$\beta_i^{women} = \gamma_0$$

$$\beta_i^{men} = \gamma_0 + \gamma_1$$

$$\beta_i^{post,men} = \text{E}(\gamma_0) + \text{E}(\gamma_1) + c_{men}(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

$$\beta_i^{post,women} = \text{E}(\gamma_0) + c_{women}(\gamma_0 + \bar{x}\gamma_1 - (\text{E}(\gamma_0) + \bar{x}\text{E}(\gamma_1)))$$

and

$$W_{women} = \beta_i^{women} - \beta_i^{post,women}$$

$$= (\gamma_0 - \text{E}(\gamma_0)) - C_{men}\left[(\gamma_0 - \text{E}(\gamma_0)) + \bar{x}(\gamma_1 - \text{E}(\gamma_1))\right]$$

$$= (1 - c_{women})(\gamma_0 - \text{E}(\gamma_0)) - C_{women}\bar{x}(\gamma_1 - \text{E}(\gamma_1))$$

$$W_{men} = \beta_i^{men} - \beta_i^{post,men}$$

$$= (\gamma_0 - \text{E}(\gamma_0)) + (\gamma_1 - \text{E}(\gamma_1) - C_{men}\left[(\gamma_0 - \text{E}(\gamma_0)) + \bar{x}(\gamma_1 - \text{E}(\gamma_1))\right]$$

$$= (1 - c_{men})(\gamma_0 - \text{E}(\gamma_0)) + (1 - C_{men}\bar{x})(\gamma_1 - \text{E}(\gamma_1))$$

Squaring

$$W_{women}^2 = (1 - c_{wom})^2 \text{Var}(\gamma_0) - 2(1 - c_{wom})c_{wom}\bar{x}\text{Cov}(\gamma_0, \gamma_1) + c_{wom}^2\bar{x}^2\text{Var}(\gamma_1)$$

$$= \frac{\left(\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\text{Var}(\gamma_0) -$$

$$2\left(\frac{[\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)][\text{Var}(\gamma_0) + \bar{x}\text{Cov}(\gamma_0, \gamma_1)]}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\right)\bar{x}\text{Cov}(\gamma_0, \gamma_1) +$$

$$\frac{\left(\text{Var}(\gamma_0) + \bar{x}\text{Cov}(\gamma_0, \gamma_1)\right)^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\bar{x}^2\text{Var}(\gamma_1)$$

$$= \frac{\bar{x}^2\text{Cov}(\gamma_0, \gamma_1)^2 + 2\bar{x}^3\text{Cov}(\gamma_0, \gamma_1)\text{Var}(\gamma_1) + \bar{x}^4\text{Var}(\gamma_1)^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\text{Var}(\gamma_0)$$

$$- 2\left(\frac{\bar{x}\text{Cov}(\gamma_0, \gamma_1)\text{Var}(\gamma_0) + \bar{x}^2\text{Cov}(\gamma_0, \gamma_1)^2 + \bar{x}^2\text{Var}(\gamma_0)\text{Var}(\gamma_1) + \bar{x}^3\text{Var}(\gamma_1)\text{Cov}(\gamma_0, \gamma_1)}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\right)\bar{x}\text{Cov}(\gamma_0, \gamma_1$$

$$+ \frac{\text{Var}(\gamma_0)^2 + 2\bar{x}\text{Var}(\gamma_0)\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Cov}(\gamma_0, \gamma_1)^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\bar{x}^2\text{Var}(\gamma_1)$$

$$= \frac{\bar{x}^2\left[\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Cov}(\gamma_0, \gamma_1)^2\right]}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)\bar{x}^2 + 2\text{Cov}(\gamma_0, \gamma_1)\bar{x}}$$

$$W_{men}^2 = (1 - c_{men})^2\text{Var}(\gamma_0) + 2(1 - c_{men})(1 - c_{men}\bar{x})\text{Cov}(\gamma_0, \gamma_1) + (1 - c_{men}\bar{x})^2\text{Var}(\gamma_1)$$

$$= \frac{\left((\bar{x} - 1)\text{Cov}(\gamma_0, \gamma_1) + (\bar{x}^2 - \bar{x})\text{Var}(\gamma_1)\right)^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\text{Var}(\gamma_0)$$

$$+ 2\frac{[(\bar{x} - 1)\text{Cov}(\gamma_0, \gamma_1) + (\bar{x}^2 - \bar{x})\text{Var}(\gamma_1)][(1 - \bar{x})\text{Var}(\gamma_0) + (\bar{x} - \bar{x}^2)\text{Cov}(\gamma_0, \gamma_1)]}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\text{Cov}(\gamma_0, \gamma_1)$$

$$+ \frac{[(1 - \bar{x})\text{Var}(\gamma_0) + (\bar{x} - \bar{x}^2)\text{Cov}(\gamma_0, \gamma_1)]^2}{\left(\text{Var}(\gamma_0) + 2\bar{x}\text{Cov}(\gamma_0, \gamma_1) + \bar{x}^2\text{Var}(\gamma_1)\right)^2}\text{Var}(\gamma_1)$$

$$= \frac{(\bar{x} - 1)^2\left[\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Cov}(\gamma_0, \gamma_1)^2\right]}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)\bar{x}^2 + 2\text{Cov}(\gamma_0, \gamma_1)\bar{x}}$$

We can see then that if $\text{Cov}(\gamma_0, \gamma_1) = 0$ we recover our result from version 2.
And we get a mean square error of

$$MSE = \left(\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Cov}(\gamma_0, \gamma_1)^2\right)\frac{(1 - p)\bar{x}^2 + p(\bar{x} - 1)^2}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)\bar{x}^2 + 2\text{Cov}(\gamma_0, \gamma_1)\bar{x}}$$

$$\frac{d}{dx}\left(\mathrm{Var}(\gamma_0)\mathrm{Var}(\gamma_1) - \mathrm{Cov}(\gamma_0,\gamma_1)^2\right)\frac{(1-p)\bar{x}^2 + p(\bar{x}-1)^2}{\mathrm{Var}(\gamma_0) + \mathrm{Var}(\gamma_1)\bar{x}^2 + 2\bar{x}\mathrm{Cov}(\gamma_0,\gamma_1)}$$

$$= 2\left(\mathrm{Var}(\gamma_0)\mathrm{Var}(\gamma_1) - \mathrm{Cov}(\gamma_0,\gamma_1)^2\right)\frac{(\mathrm{Var}(\gamma_0)(\bar{x}-p) + \mathrm{Var}(\gamma_1)p(\bar{x}-1)\bar{x} + \mathrm{Cov}(\gamma_0,\gamma_1)(\bar{x}^2-p))}{(\mathrm{Var}(\gamma_0) + \bar{x}(\mathrm{Var}(\gamma_1)\bar{x} + 2\mathrm{Cov}(\gamma_0,\gamma_1)))^2}$$

Thus we can use the quadratic formula to determine the optimal trial proportions.

$$\bar{x} = \frac{\mathrm{Var}(\gamma_1)p - \mathrm{Var}(\gamma_0) \pm \sqrt{4p\left[\mathrm{Var}(\gamma_0) + \mathrm{Cov}(\gamma_0,\gamma_1)\right]\left[\mathrm{Var}(\gamma_1)p + \mathrm{Cov}(\gamma_0,\gamma_1)\right] + \left(\mathrm{Var}(\gamma_0) - \mathrm{Var}(\gamma_1)p\right)^2}}{2(\mathrm{Var}(\gamma_1)p + \mathrm{Cov}(\gamma_0,\gamma_1))}$$

and clearly $\mathrm{Var}(\gamma_1)p + \mathrm{Cov}(\gamma_0,\gamma_1) \neq 0$

We can thus see that we recover our original case of choosing between all men and all women whenever

$$\mathrm{Cov}(\gamma_0,\gamma_1) = -\mathrm{Var}(\gamma_0)$$

And when this holds, when $p = \frac{\mathrm{Var}(\gamma_0)}{\mathrm{Var}(\gamma_1)}$, we again get our edge-case where the the error is independent of the trial proportions and any $\bar{x}$ produces an equal error

Now we can check whether our setup in version 1 satisfies this.

Recall we have

$$\beta_i = \gamma_0 + x(\gamma_1 - \gamma_0)$$

Let's call $\gamma_2 = \gamma_1 - \gamma_0$

First observe that

$$\mathrm{Var}(\gamma_2) = \mathrm{Var}(\gamma_1 - \gamma_0) = \mathrm{Var}(\gamma_0) + \mathrm{Var}(\gamma_1) \quad \text{assuming both are 1}$$

$$= 1 + 1 = 2$$

and

$$\mathrm{Cov}(\gamma_0,\gamma_2) = \mathrm{Cov}(\gamma_0,\gamma_1 - \gamma_0) = \mathrm{Cov}(\gamma_0,\gamma_1) - \mathrm{Cov}(\gamma_0,\gamma_0)$$

$$= 0 - 1 = -1$$

And so we recover that

$$\mathrm{Cov}(\gamma_0,\gamma_2) = -\mathrm{Var}(\gamma_0)$$

And more over that

$$p = \frac{\mathrm{Var}(\gamma_0)}{\mathrm{Var}(\gamma_2)} = \frac{1}{2}$$

And thus and MSE

$$MSE = \left(\text{Var}(\gamma_0)\text{Var}(\gamma_1) - \text{Cov}(\gamma_0, \gamma_1)^2\right) \frac{(1-p)\bar{x}^2 + p(\bar{x}-1)^2}{\text{Var}(\gamma_0) + \text{Var}(\gamma_1)\bar{x}^2 + 2\text{Cov}(\gamma_0, \gamma_1)\bar{x}}$$

$$= (2-1)\frac{\frac{1}{2}\bar{x}^2 + \frac{1}{2}(\bar{x}^2 - 2\bar{x} + 1)}{1 - 2\bar{x} + 2\bar{x}^2}$$

$$= \frac{\frac{1}{2}\bar{x}^2 + \frac{1}{2}(\bar{x}^2 - \bar{x} + 1)}{2(\frac{1}{2} - \bar{x} + \bar{x}^2)} = \frac{1}{2}$$

Which is always equal to $\frac{1}{2}$ and independent of $\bar{x}$
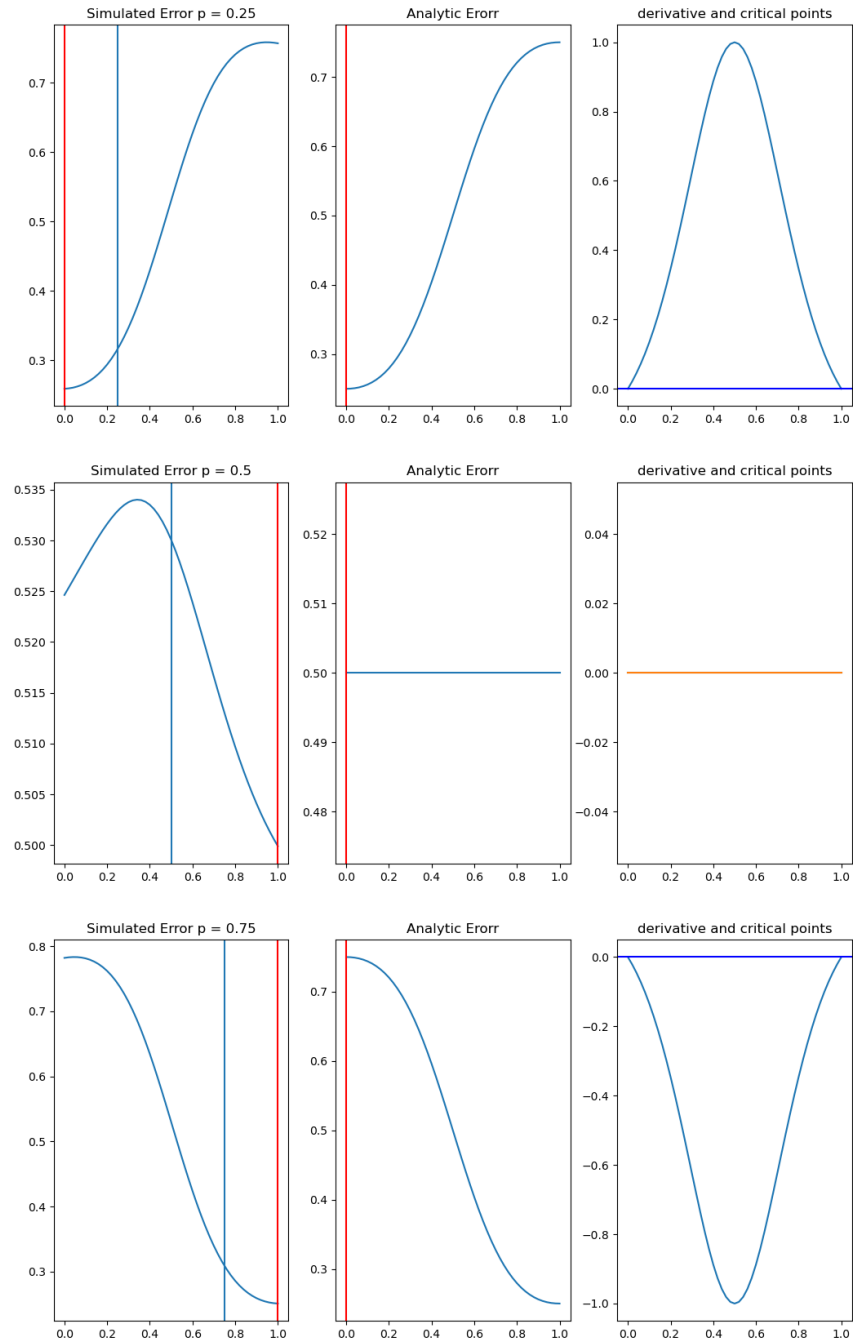
Figure 1: First column contains the simulated error for different population proportions with the population proportion shown by the blue vertical line and point of minimum error shown with the red vertical line. The second column has the analytic error computed in the math above as well as a vertical line showing the first critical point. The third column shows the first derivative with respect to $\bar{x}$.
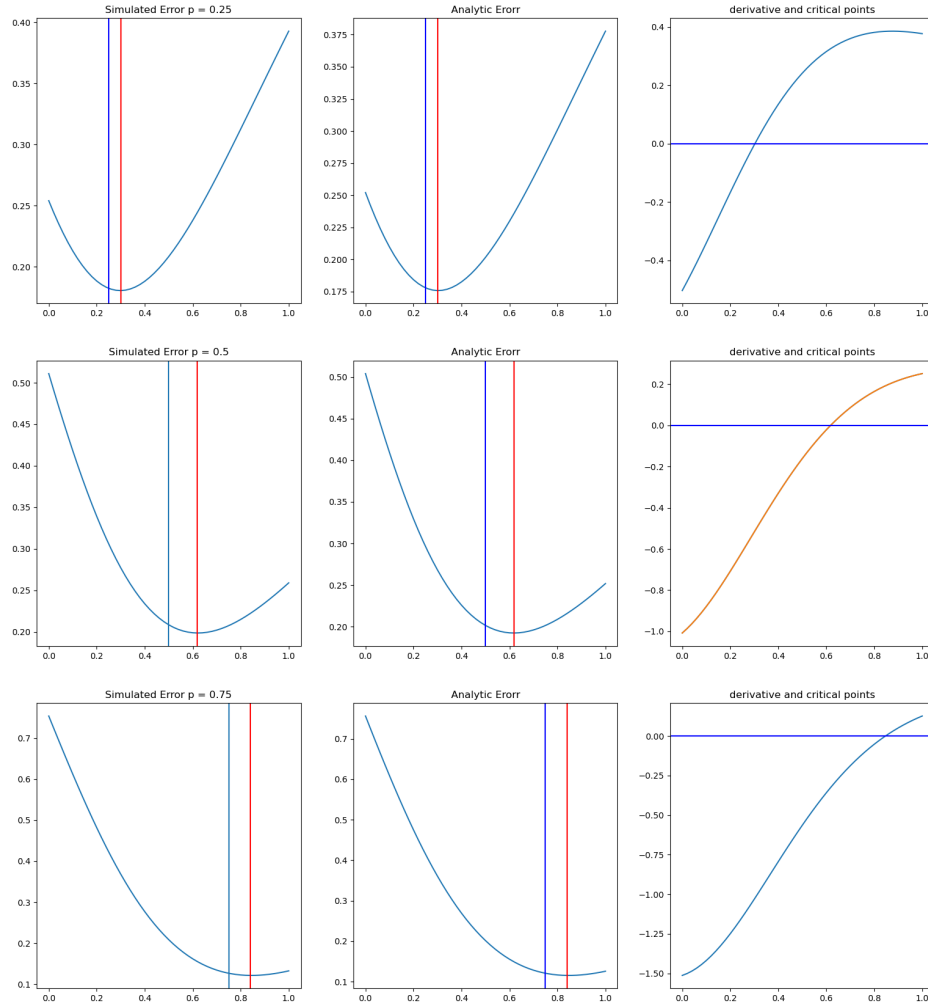
Figure 2: First column contains the simulated error for different population proportions with the population proportion shown by the blue vertical line and point of minimum error shown with the red vertical line. The second column has the analytic error computed in the math above as well as a vertical line showing the first critical point and blue line showing population proportion. The third column shows the first derivative with respect to $\bar{x}$ and where it equals zero.
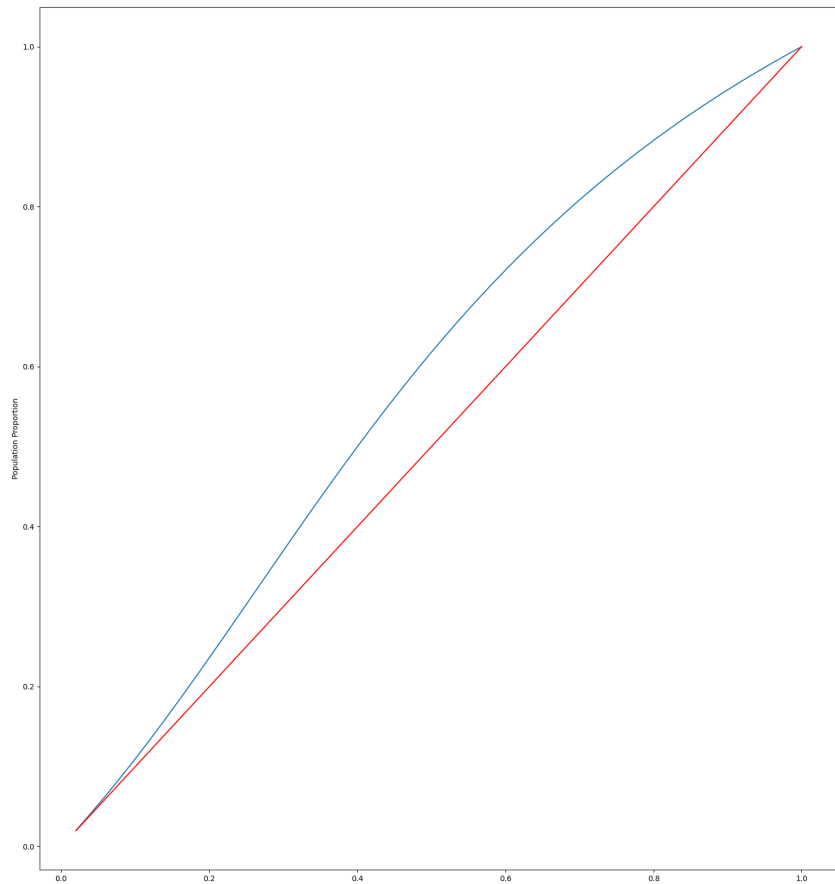
Figure 3: Here we show the optimal trial proportion of men vs the population proprtion (on the y and x axis respectively) for equal variances of 1.