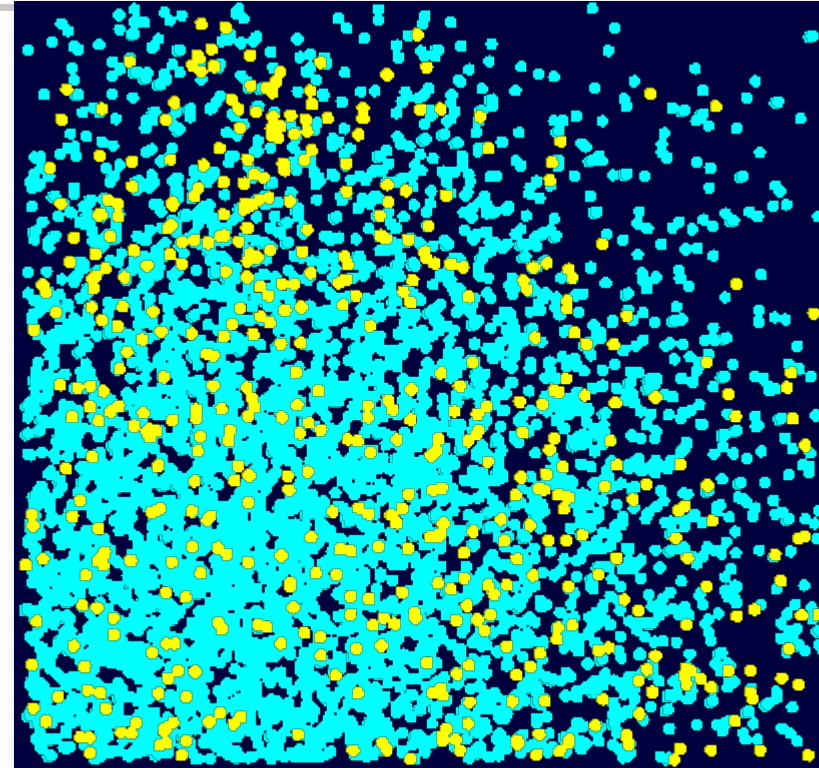# Using Prior Probabilities and Profit/Loss

Outcome Overrepresentation

A common predictive modeling practice is to build models from a sample with a primary outcome proportion that is **different** from the original population.
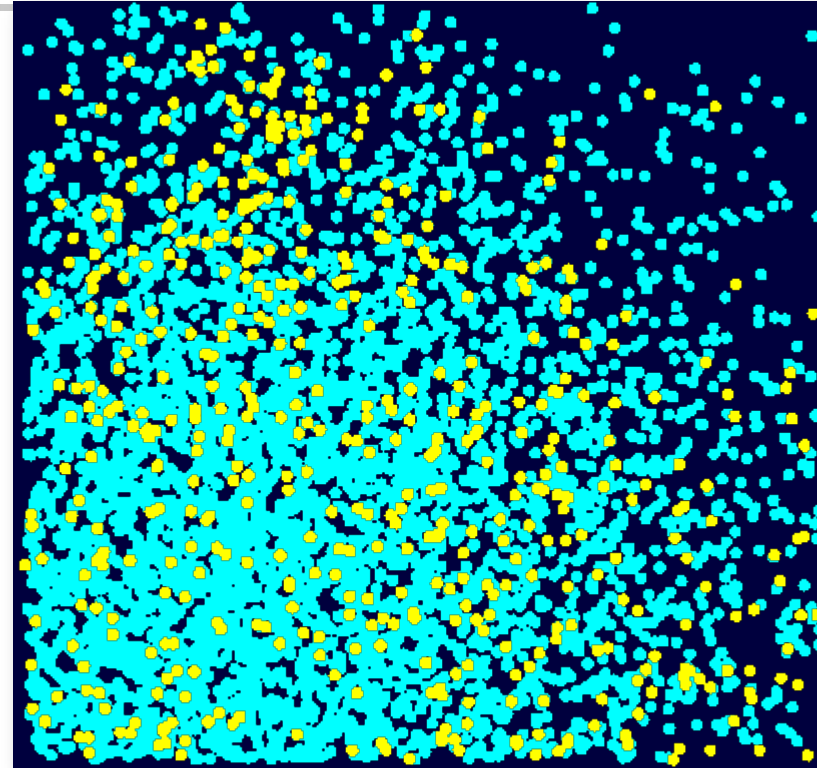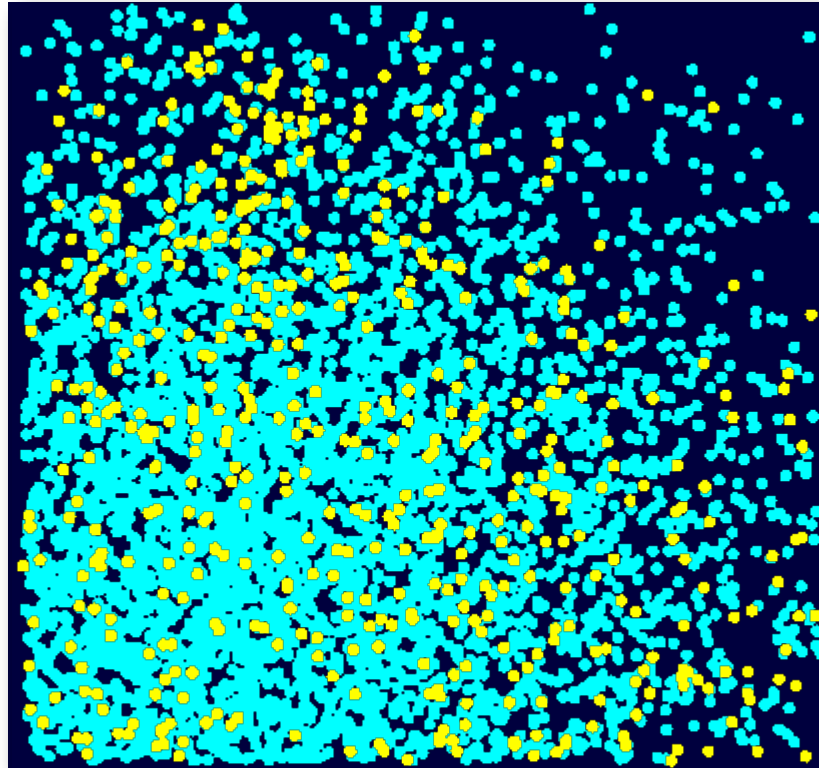
Outcome Overrepresentation

A common predictive modeling practice is to build models from a sample with a primary outcome proportion that is different from the original population.
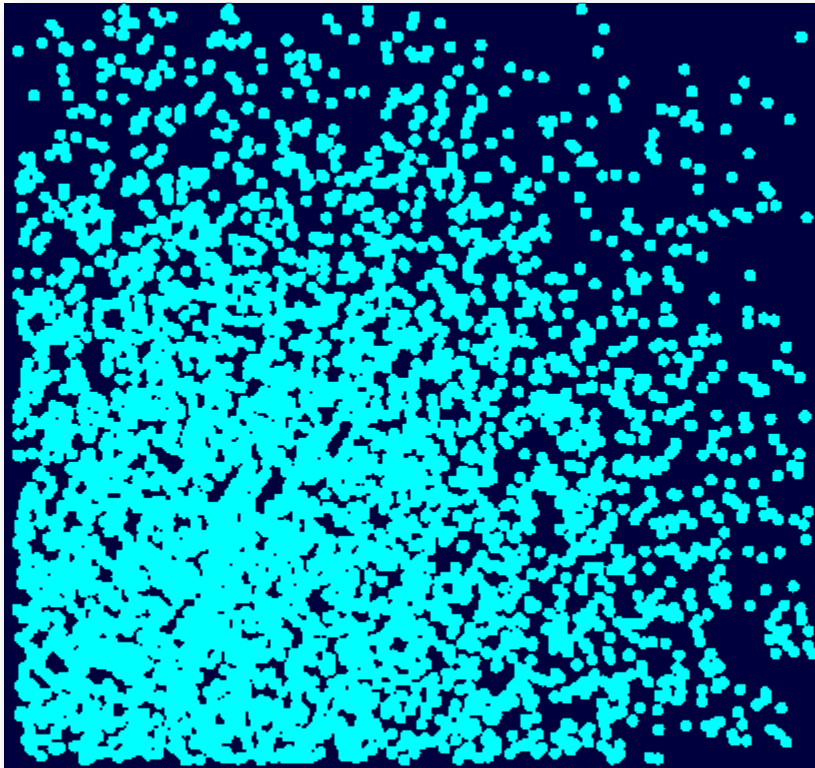
Separate Sampling

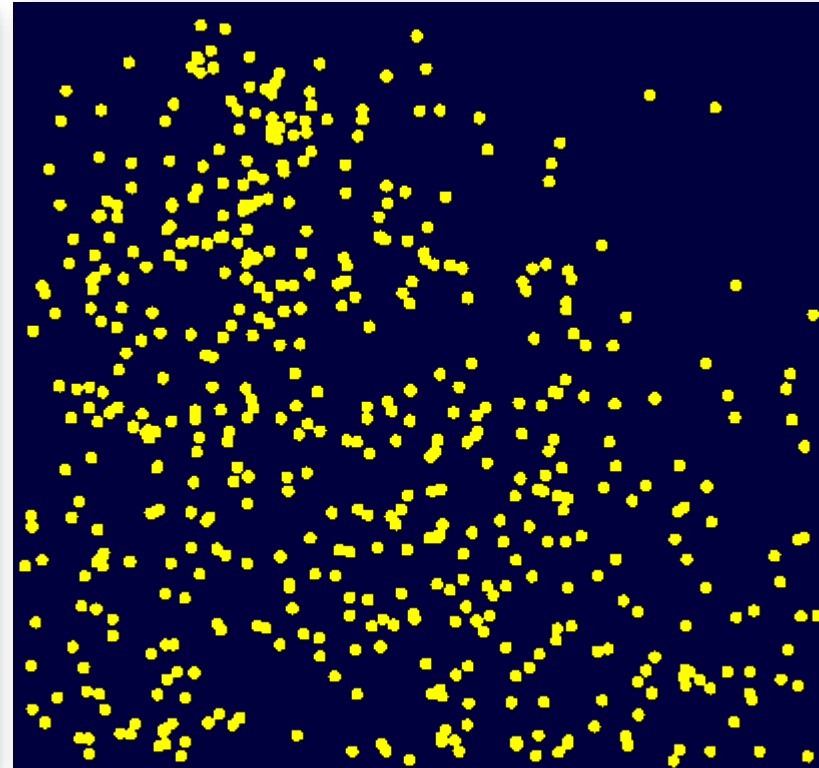*secondary outcome*   *primary outcome*

Target-based samples are created by considering the primary outcome cases separately from the secondary outcome cases.

...

Separate Sampling
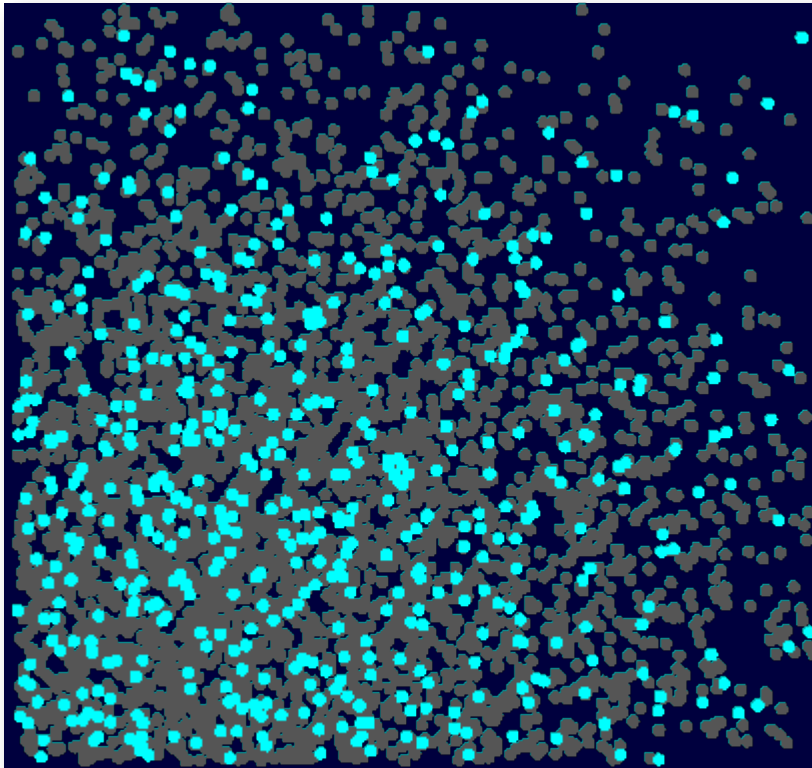

secondary outcome   *primary outcome*

Target-based samples are created by considering the primary outcome cases separately from the secondary outcome cases.
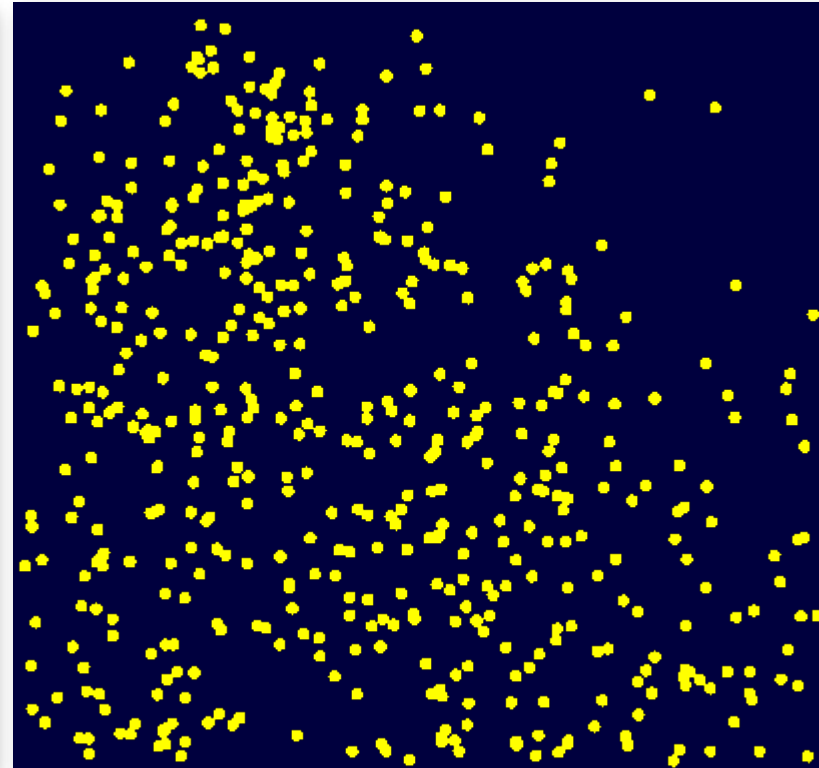
...

Separate Sampling

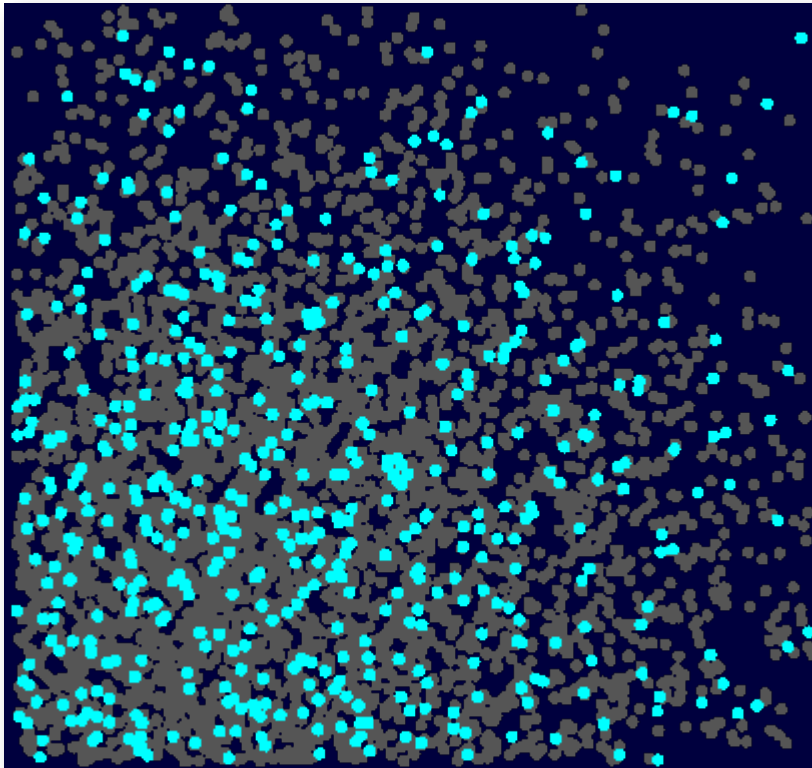*secondary outcome*

primary outcome

Select some cases.

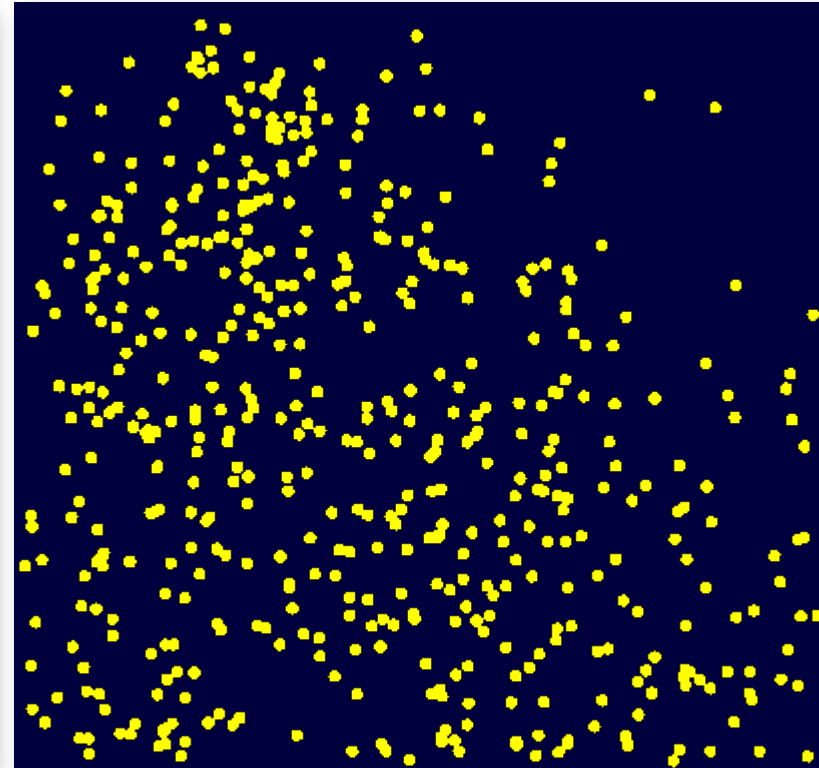Select all cases.

...

Separate Sampling

secondary outcome
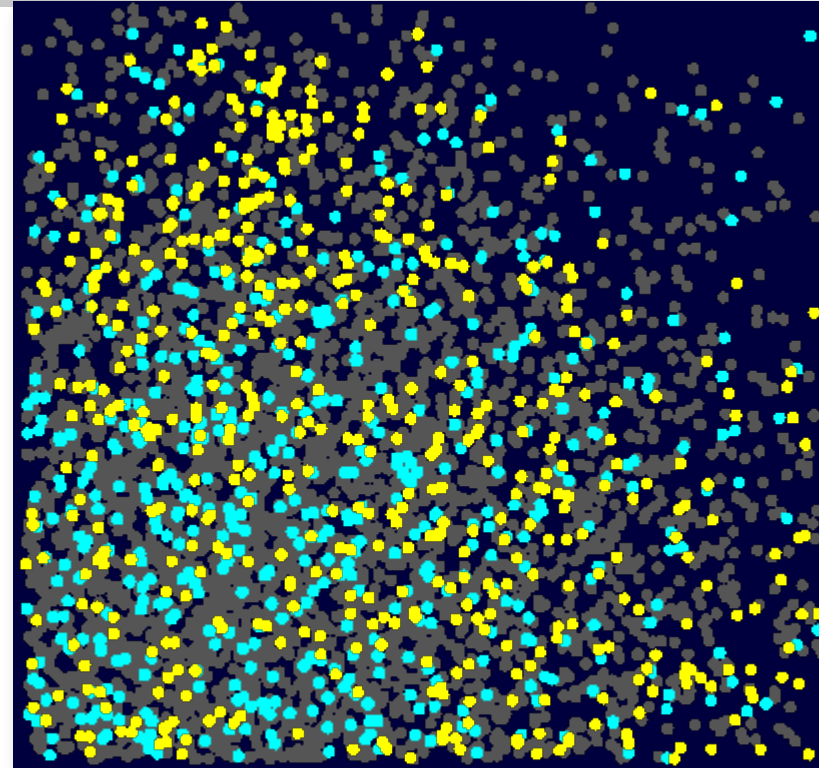

Select some cases.

primary outcome

Select all cases.

...

The Modeling Sample

- + Similar predictive power with smaller case count

- − Must adjust assessment statistics and graphics

- − Must adjust prediction estimates for bias

Demo: Adjusting for Prior Probabilities

- Start with the Diagram3_BAN5743 (does not have a metadata node)
  - Add the data set PVA97NK_Copy to your data sources
  - Make sure that for all of the variables, their roles and levels match what is expected in this data (as described earlier)
- Select the PVA97NK_copy node.
- Click the Decisions property ellipsis button
- Click the **Prior Probabilities** tab
- Select **Yes**. The dialog box is updated to show the Adjusted Prior column
- Enter **0.05** as the Adjusted Prior value for the primary outcome, Level 1. Enter **0.95** as the Adjusted Prior value for the secondary outcome, Level 0. Select **OK** to close the Decision Processing dialog box
- For the Neural Node, select **Model Selection Criterion** ⇨ **Average Error**
- Click the **Model Comparison** node and examine the Properties panel.
- Under the Model Selection portion of the Properties pane, select **Average Squared Error** for **Selection Statistic**, and set the **Selection Table** to **Validation**.
- Run Model Selection and examine results

Profit Matrices

profit distribution
for solicit decision
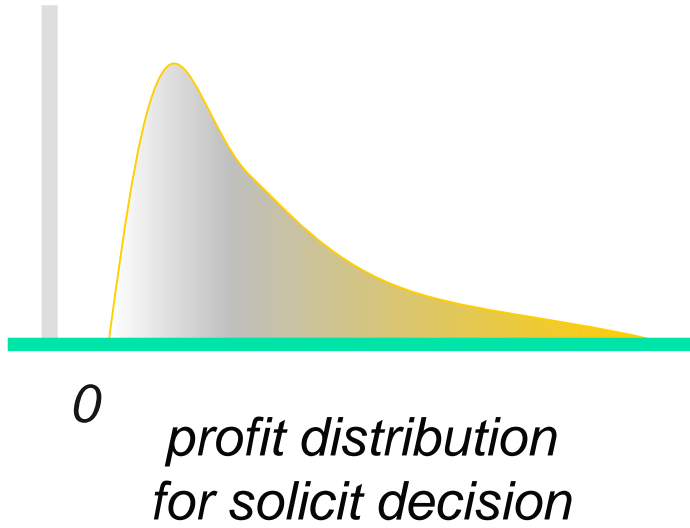
	solicit	ignore
primary outcome	**15.14**	*0*
secondary outcome	*-0.68*	*0*

Profit Matrices

	solicit	ignore
primary outcome	15.14	0
secondary outcome	-0.68	0

0

profit distribution for solicit decision
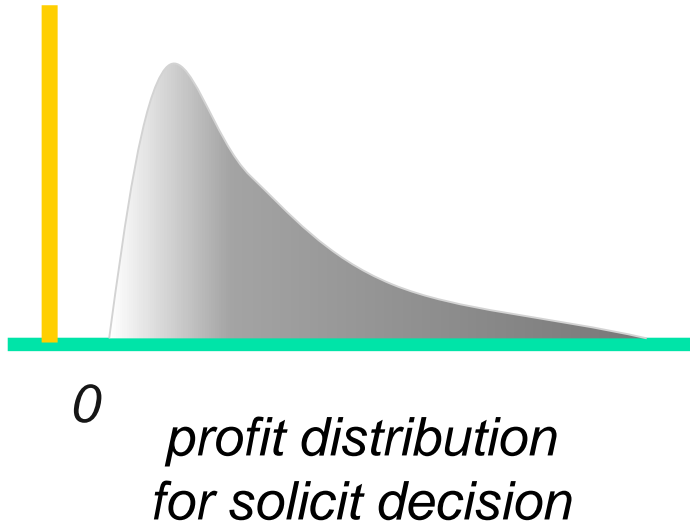
Decision Expected Profits
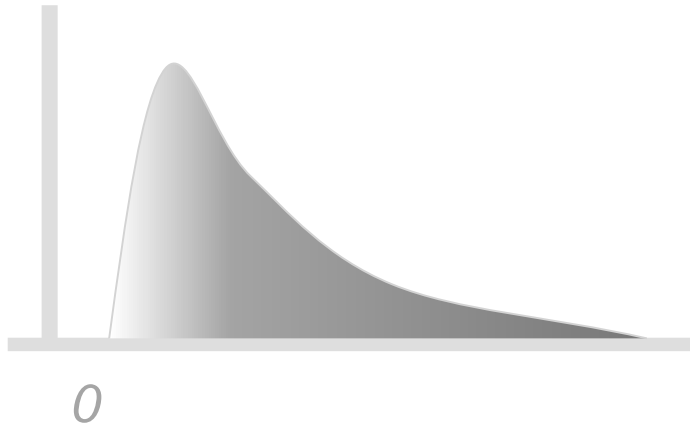
	solicit	*ignore*
primary outcome	**15.14**	**0**
secondary outcome	**-0.68**	**0**

0

Choose the larger.

Expected Profit Solicit = 15.14 $\hat{p}_1$ − 0.68 $\hat{p}_0$

Expected Profit Ignore = 0

...

Decision Threshold

	solicit	ignore
primary outcome	**15.14**	**0**
secondary outcome	**-0.68**	**0**

0
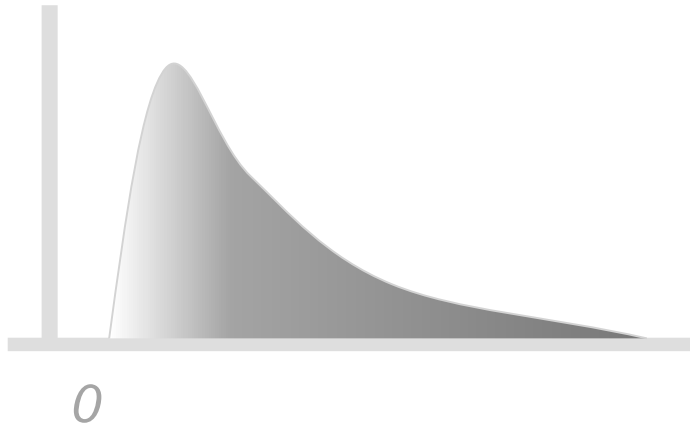
decision threshold

$\hat{p}_1 \geq 0.68 / 15.82 \Rightarrow$ **Solicit**

$\hat{p}_1 < 0.68 / 15.82 \Rightarrow$ **Ignore**

Average Profit

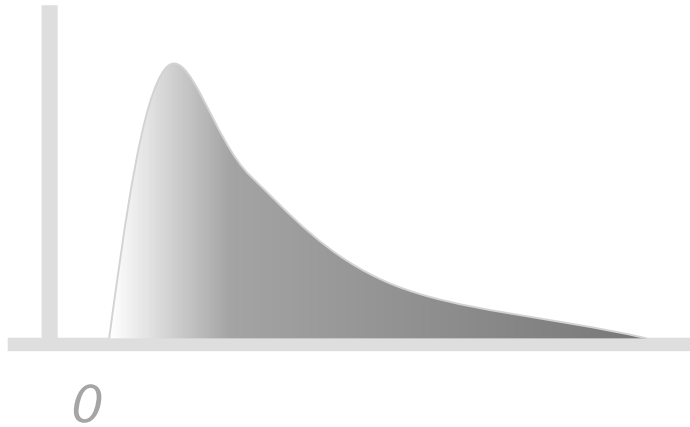|	*solicit*	*ignore*
primary outcome	**15.14**	**0**
secondary outcome	**-0.68**	**0**

0

average profit

Average profit = (15.14 × N_{PS} − 0.68 × N_{SS}) / N

N_{PS} = # solicited primary outcome cases

N_{SS} = # solicited secondary outcome cases

N = total number of assessment cases

Demo: Adjusting for Prior Probabilities

- Add a StatExplore node to Data partition node. Right-click > Edit variables > Change Use role of TargetD to Yes. Run StaExplore node to check for average of TargetD

- Select the PVA97NK_copy node. Click the Decisions property ellipsis button

- Click the Decisions tab. Select Yes to use the decisions

- Enter Solicit (to replace the word DECISION1) in the first row of the Decision Name column.

- Enter Ignore (to replace the word DECISION2) in the second row of the Decision Name column. (Press Enter to make sure that the decision names are saved.)

- Click the Decision Weights tab. Enter the profit values (15.14 and -0.68 under Solicit and 0 and 0 under ignore) into the corresponding cell of the profit weight matrix. Press **Enter** to make sure that the changes are saved.)

- Select **OK** to close the Decision Processing dialog box

- 18.   Change the Selection Statistic in the Properties panel of the Model Comparison node to Average Profit/Loss.

- Run Model Selection and examine results

Self Study Handout

- After you define a profit matrix, it is possible to optimize your model strictly on profit. Instead of seeking the model with the best prediction estimates, you find the model with best prediction decisions (those that maximize expected profit).

- The Decision Tree models don't need any changes for doing profit maximization

- Select the **Regression (optimal)** node.

- Select **Selection Criterion** ⇨ **Validation Profit/Loss**.

- Select the **Neural Network** node.

- Select **Model Selection Criterion** ⇨ **Profit/Loss**

- Run the Model Comparison node and view the results.