

In-Class Exercise: Calculating and Interpreting Chi-Squared Statistics

Remember the logworth value, the metric that determines a split in a decision tree algorithm, is the log of the p-value derived from the Chi-Squared statistic. So the logworth value is directly related to a Chi-Squared calculation comparing observed outcomes versus expected outcomes.

With this in mind, consider these two hypothetical scenarios for a different set of donor data:

Scenario 1: Gender as a split variable

	Observed		
	Male	Female	
Gift	375	125	500
No Gift	425	75	500
	800	200	1000

	Expected		
	Male	Female	
Gift	400	100	500
No Gift	400	100	500
	800	200	1000

Scenario 2: Age as a split variable

	Observed		
	<40	>=40	
Gift	405	95	500
No Gift	395	105	500
	800	200	1000

	Expected		
	<40	>=40	
Gift	400	100	500
No Gift	400	100	500
	800	200	1000

1) What is the Chi-Squared value for Scenario 1 and Scenario 2?

Scenario 1 (Gender): _____

Scenario 2 (Age): _____

2) Which split variable, Gender or Age, is a more powerful differentiator whether someone will donate a gift?

Answer: _____

3) Briefly explain why:

Answer: _____

4) Which split variable, Gender or Age, will have the higher logworth value?

Answer: _____