

DSC 433/533 – Class 2 – Charles Book Club Example

Task: Find out whether males or females are more likely to buy “The Art History of Florence.” How much profit does the company expect to make by mailing to everyone (if it costs \$0.50 to make an offer, we receive \$6 if customer purchases book, and we have 500,000 customers in our database)? Only males? Only females?

Sample mailing: 4000 customers (1800 *training*, 1400 to *validation*, 800 *test*).

Expected profit if mail everyone in validation sample = $122 * 6 - 1400 * 0.5 = \$32$.

Scale-up to 500,000 customers in database = $(500,000/1400) * 32 = \$11,429$.

Can we do better by mailing only to males or only to females?

Use Excel PivotTable to find out # males and females in training sample purchasing book:

Sum of Florence	Gender		Grand Total
	0	1	
Total	50	101	151

Count of Florence	Gender		Grand Total
	0	1	
Total	545	1255	1800

In other words, $50/545 = 9.2\%$ of males purchased the book, but only $101/1255 = 8.0\%$ of females purchased the book. So, pick males.

What happens in the validation sample?

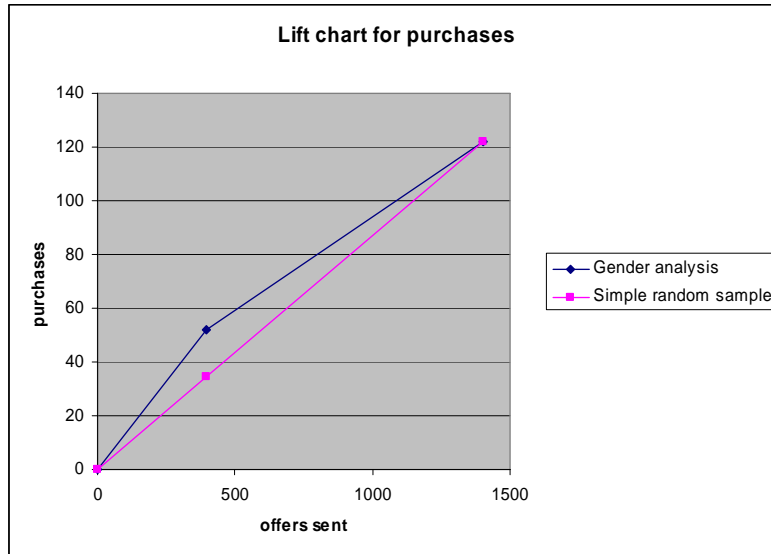
Sum of Florence	Gender		Grand Total
	0	1	
Total	52	70	122

Count of Florence	Gender		Grand Total
	0	1	
Total	394	1006	1400

Lift chart for purchases:

Cumulative offers sent	0	394	1400
Gender analysis	0	52	122
Simple random sample	0	34.3	122

Random sample calculation: $394 * 122/1400 = 34.3$.

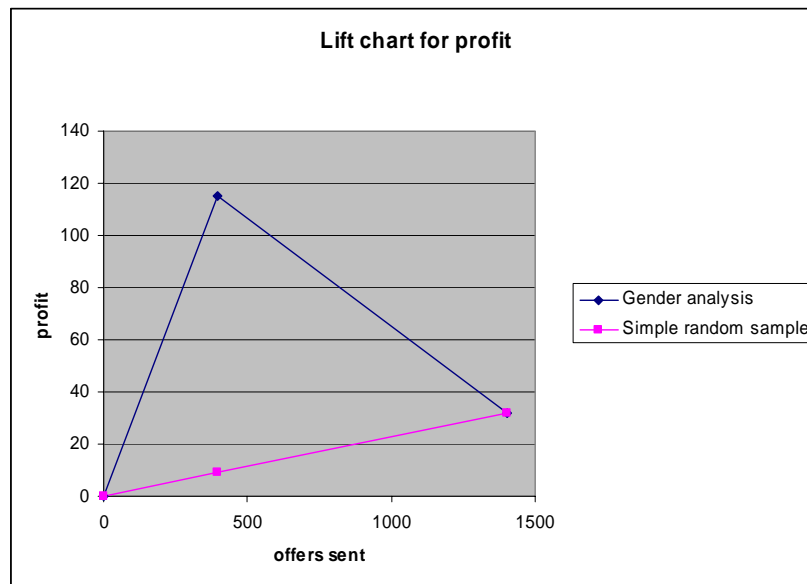


Lift chart for profit:

Cumulative offers sent	0	394	1400
Gender analysis	0	115	32
Simple random sample	0	9	32

Gender analysis profit calculation: $52 \cdot 6 - 394 \cdot 0.5 = 115$.

Random sample profit calculation: $34.33 \cdot 6 - 394 \cdot 0.5 = 9$.



52/394 males purchased the book.

Expected profit if mail only males in validation sample = $52 \cdot 6 - 394 \cdot 0.5 = \115 .

Scale-up to 500,000 customers in database = $(500,000/1400) \cdot 115 = \$41,071$.

Alternative calculation since we expect $500,000 \cdot 394/1400 = 140,714.29$ males is $(140,714.29/394) \cdot 115 = \$41,071$.