Exploration of Magazine Advertising Costs Using Supervised Learning

Andy W. Chen
University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T 1Z2, Canada

Abstract
Magazine advertising is a lucrative business and advertisers must learn about the potential impact of their efforts and the various factors that will affect the advertising outcome. In this paper, I explore the relationship between the cost of advertising in magazines, magazine circulation, and demographics of the readers. I use supervised machine learning methods to build models that will predict the cost of advertising given the factors mentioned above. I conduct F-tests and find that magazine circulation and median income of readers have significant predictive power when predicting advertising cost in magazines. The model has an R2 of 0.6492, suggesting that the factors in the model explain around 64.92% of the variation in the advertising cost of magazines.

Keywords
Magazine Advertising, Marketing, Supervised Learning, Business Management, Data Science

I. Introduction

The cost of a page of advertising in a magazine can be attributed to a number of factors including but not limited to its popularity, its audience's demographics, its scale of operations and probably something relating to the value of the ad to the advertiser. The purpose of this paper is to explore the impact of a magazine's circulation, its percentage of male readership and the median income on the cost of a one-page advertisement in a magazine. I use a supervised machine learning model, the multiple linear regression to model the relationship.

An overall F-test shows that at least one of the predictor variables has explanatory power over the page cost, whereas partial t-tests showed that the data only contained enough evidence to conclude that circulation and median income have individual explanatory power over page cost. The R^2 of this model is 0.6492 and suggests that 64.92% of the variability in page cost can be explained by our model.

Related works on advertising in magazines include one by Zeisel and Harper[1], who find statistically significant differences between ratings for identical advertisements in different magazines and that these differences are caused by the number of volumes produced by each magazine. Soley and Krishnan[2] find no evidence that advertising decreases the price per page or the price per copy of consumer magazines. Olson[3] study the portrayal of advertisements in magazines published between 1900 and 1940. He finds that articles that mentioned advertising are most common in elite and specialized magazines. Depken II and Wilson[4] discover that advertising used to replace editorial content reduces the quality and price of magazine subscriptions. Rotfeld and Parsons[5] find that publishers would use personal review on an ad hoc basis to determine the suitability of particular advertisements.

II. Methodology

I use a dataset comprised of unit page cost for color advertisements, magazine circulation, percentage of male readers of the magazine, and median income of the readers. This dataset contains data for the above variables for 48 well-known magazines such as Business Week, Cosmopolitan, Forbes, Time, and Sports Illustrated. The data can be found in Table 1.

I use a supervised machine learning model, the linear regression, to estimate the effect of magazine circulation, percent male, and median income on the average unit cost per page of advertising. The idea behind this approach is that the price that a magazine charges for printing an advertisement is highly variable across different magazines. The price charged is based on its ability to get that price from potential advertisers. The various reasons for this variability in prices could be the popularity of the magazine, its target audience, its scale of operations and probably something relating to the value of the ad to the advertiser. In order to estimate the effect of different factors on advertising costs, I build a model that will give quantitative measurements of these effects. The proposed model is a multiple linear regression model of the form

$$\text{Page Cost} = \beta_1 \text{Circulation} + \beta_2 \text{Percent Male} + \beta_3 \text{Median Income} + \varepsilon$$

Estimating this model for the given dataset will give us the three parameters which give the estimates for the average change in page cost associated with every one thousand person change in circulation, with one percent change in male readership and with each one dollar change in median income respectively for given levels of the other variables. I also validate the model by testing the assumptions underlying it. If the assumptions are satisfied, the results in the model can be viewed as being true and reasonable.

The findings from this model will help determine whether circulation, percentage of male readership and median income of the audience have any significant effect on the cost of an advertisement page or not. I will test the overall significance of the model using the F-test and R^2. This will help understand advertising costs in magazines and their drivers. I am also interested in determining which factor explains the most variability in the response variable, unit page cost.

Table 1 : Data used for the supervised learning model

Magazine	Page Costs	Circulation (1000s)	% Male	Median Income
Better Homes & Gardens	$73,820	8,000	22	$23,241
Business Week	$35,140	845	72	$30,884
Car & Driver	$23,795	725	88	$25,982
Cosmopolitan	$28,980	2,250	17	$22,785
Ebony	$21,886	1,250	42	$16,505

ISSN : 2348-6503 (Online)
ISSN : 2348-893X (Print)

Family Circle	$62,750	7,450	13	$21,785
Field & Stream	$33,760	2,000	79	$24,337
Forbes	$25,090	700	74	$36,783
Fortune	$30,040	670	71	$35,204
Glamour	$24,340	1,800	6	$21,828
Golf Digest	$26,625	1,025	82	$32,949
Good Housekeeping	$58,020	5,000	13	$21,980
Harper's Bazaar	$16,200	650	8	$25,358
Hot Rod	$20,400	850	78	$23,056
House & Garden	$25,430	1,000	19	$23,726
House Beautiful	$19,775	800	16	$24,198
Ladies' Home Journal	$48,000	5,000	12	$21,583
Mademoiselle	$16,280	1,000	9	$23,660
McCalls	$59,830	6,200	12	$20,690
Mechanix Illustrated	$24,815	1,600	81	$22,568
Money	$25,740	1,000	58	$31,587
Motor Trend	$21,905	750	84	$23,878
Ms.	$7,845	450	16	$24,107
National Enquirer	$26,500	4,637	40	$19,969
National Geographic	$95,575	8,400	55	$26,294
National Lampoon	$9,900	580	72	$22,888
Newsweek	$63,850	2,950	62	$26,719
Outdoor Life	$28,475	1,500	82	$23,596
Parents Magazine	$36,960	1,650	22	$20,779
People	$56,425	2,350	40	$23,971
Playboy	$55,710	5,000	78	$24,051
Playgirl	$7,220	650	36	$19,329
Popular Mechanics	$26,932	1,600	81	$25,474
Popular Science	$26,820	1,800	80	$26,542
Reader's Digest	$97,700	17,900	44	$21,802
Redbook	$42,675	3,800	11	$22,794
Road & Track	$18,775	630	89	$28,093
Rolling Stone	$17,770	700	66	$24,074
Scientific American	$24,000	720	68	$29,531
Seventeen	$19,250	1,500	9	$21,251
Sports Illustrated	$54,165	2,250	80	$26,275
The Star	$21,350	3,400	35	$19,156
Time	$85,870	4,400	56	$26,908
True Story	$13,435	1,400	18	$14,325
TV Guide	$77,400	17,345	45	$20,461
U.S. News and World Report	$42,510	2,050	63	$26,998
Vogue	$18,000	950	11	$23,452
Woman's Day	$60,435	7,125	8	$21,910

III. Results and Discussion

I start with some exploratory analysis on the data. Figure 1 below shows the histograms of the advertising cost per page and magazine circulation. By examining the histogram of advertising cost per page it can be observed that the distribution of advertising cost is asymmetric and is skewed to the left. The magazine circulation distribution (the right-hand graph) is also asymmetrically skewed to the left.

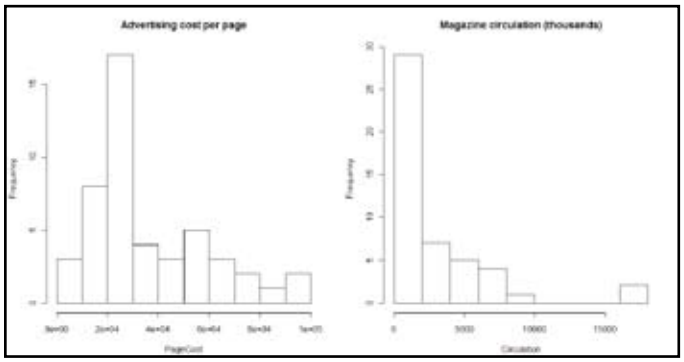

Fig. 1: Histograms of advertising cost per page & magazine circulation

Figure 2 shows the histograms of the percentage of male readership and the median income of readership. The distribution of the percentage of male readership exhibits bimodal behaviour and that the distribution of median income is nearly symmetrical about its mean, with a slight right skewness.

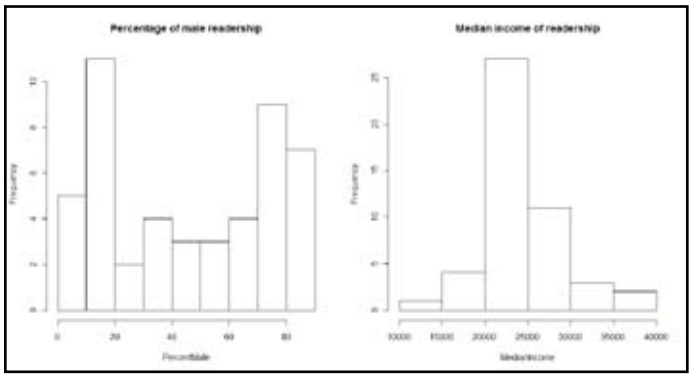

Fig. 2: Histograms of percentage of male readership & median income of readers

Figures 3 and 4 show boxplots of the variables contained in the dataset. There are some outliers shown in the boxplots for the level of magazine circulation and the median income of the readership.

The outliers in the magazine circulation boxplot correspond to Reader's Digest and TV Guide. The outliers in the median income boxplot correspond to Forbes magazine and Fortune magazine. I make note of these outliers so as to be aware of their potential for influencing our analysis. However, due to the high distribution level of Reader's Digest and TV Guide I have no reason to suspect that these data points are inaccurate. Similarly in the case of Forbes magazine and Fortune magazine; given the target audience for these magazines it is reasonable to trust that there is no inaccuracy in the data and that these values are legitimate.

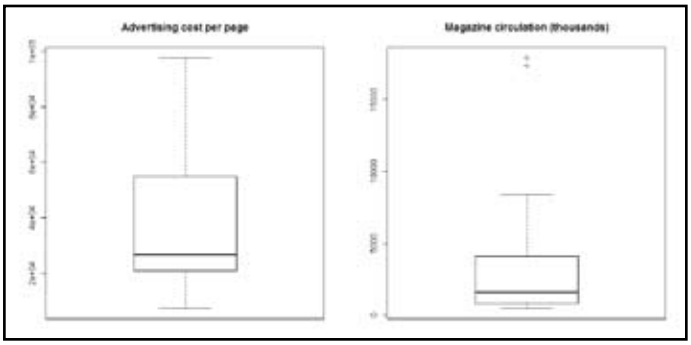

Fig. 3: Boxplots of advertising cost per page and magazine circulation (thousands)

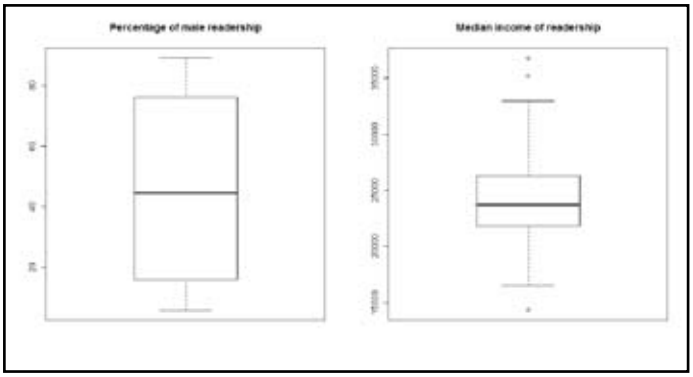

Fig. 4: Boxplots of percentage of male readership & median income of readers

The results of estimating the least squares equation is

Mean Page Cost = -8642.5 + 5.3 Circulation + 11.0 Percent Male + 1.2 Median Income

To evaluate the machine learning model, I use the overall F-test to determine the significance of the linear relationship between the response and predictor variables in the multiple linear regression model. The hypotheses to be tested are:

H_0: $\beta_1 = \beta_2 = \beta_3$

H_a: At least one of β_1, β_2, β_3 not equal 0

The test statistic that will be used to test H_0 H_0 against H_a is given by F = MSR/MSE, which turned out to be 33.3. Under H_0 H_0, the sampling distribution of the F-statistic is an F-distribution with 3 degrees of freedom in the nominator and 44 degrees of freedom in the denominator. I reject H_0 at the 5% significance level as P-value < 0.00001. Thus, the result of the test is significant at the 5% significance level. The data provide sufficient evidence that there is a linear relationship (or association) between the response variable page cost and the predictor variables circulation, percentage of male readership, and median income. The coefficient of multiple determination R^2 R^2 is 0.6492, meaning that is the 64.92% of the total variation in the dependent variable (page cost) are explained by all of the independent variables (circulation, percentage of male readership, and median income) taken together.

The least squares estimator of the regression intercept is -8642.5. This intercept is the expected value of the advertising page costs with no circulation, 0% male readership, and $0.00 median income. This is not interpretable our dataset does not include data points near the origin. The intercept, therefore, has no significant meaning physically. The slope is 5.3 for circulation. This means that each one-thousand magazine increase in circulation is associated with a $5.3 increase in average page cost for a magazine that has a particular percentage of male readership and particular median income.

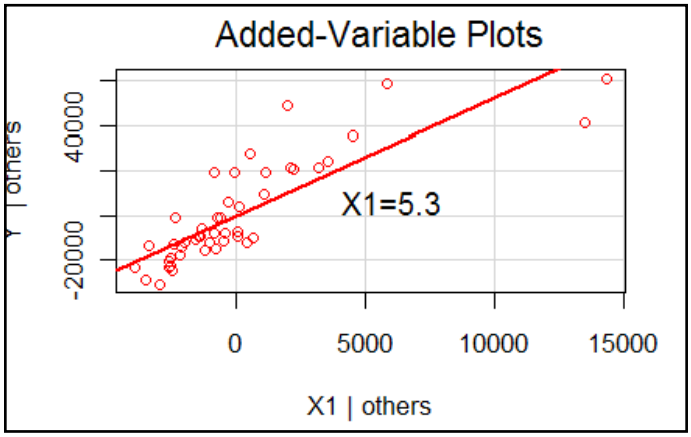

Fig. 5: Added variable plot of circulation

The slope is -11.0 for percent of male readership. This means that each one percent increase in male readership is associated with an $11.0 decrease in average page cost for a magazine that has a particular circulation and particular median readership income.

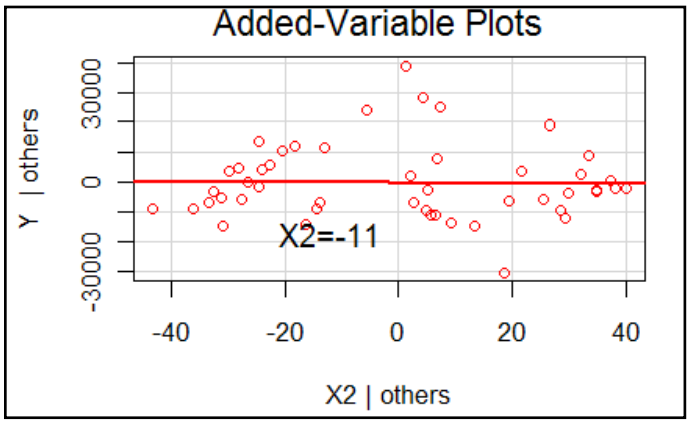

Fig. 6: Added variable plot of percentage of male readership

The slope is 1.2 for median income of readers. This means that each one dollar increase in median income is associated with a $1.2 increase in average page cost for a magazine that has a particular circulation and particular percentage of male readership.

ISSN : 2348-6503 (Online)
ISSN : 2348-893X (Print)

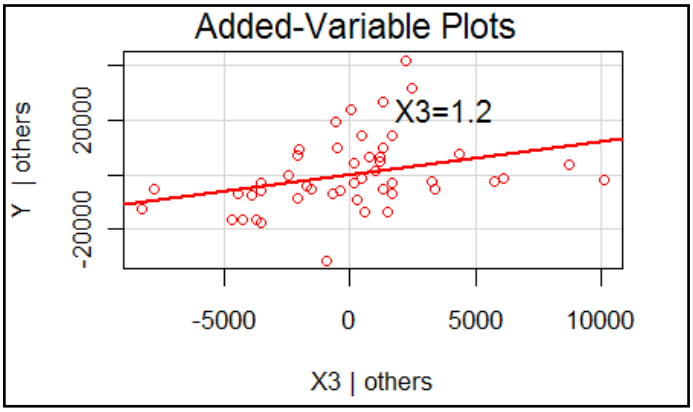

Fig. 7: Added variable plot of median income of readers

The coefficient of determination R^2 has a value of 0.6492, which indicates a moderately good fit of the model. However, improvements can certainly be made if more data about the same variables or new variables are available.

IV. Conclusions

This paper show the results of a supervised machine learning used to explore the relationship between factor that affect unit page cost of advertising and make predictions based on these features. The results are evaluated with statistical test, the F-test, and I found that the features, or predictor variables, are jointly significant. The impact of percentage of male readership has the largest magnitude and is negative, while the impact of median income of the readers has the smallest magnitude and is positive. In the future, it would be interesting to gather data on additional variables that may affect the cost of advertising such as magazine ratings, genre, publication frequency, and page number (or specific location inside the magazine).

References

[1] *Zeisel H, Harper VE. The Advertising Value of Different Magazines. Journal of Marketing. 1948;13(1):56-61.s*
[2] *Soley L, Krishnan R. Does Advertising Subsidize Consumer Magazine Prices? Journal of Advertising. 1987;16(2):4-9.*
[3] *Olson E. How Magazine Articles Portrayed Advertising from 1900 to 1940. Journal of Advertising. 1995;24(3):41-54.*
[4] *Depken II CA, Wilson DP. Is Advertising a Good or a Bad? Evidence from U.S. Magazine Subscriptions. The Journal of Business. 2004;77(S2):61-80.*
[5] *Rotfeld HJ, Parsons PR. Self-Regulation and Magazine Advertising. Journal of Advertising. 1989;18(4):33-40.*