

Machine Learning and Data Sciences

INFO 6105, Fall 2019

QUIZ 1, SATURDAY SEP 28, 2019

Answer the questions in the spaces provided on the question sheets. If you run out of space for an answer, continue on the back of the page. Write your name below and in the top right corner of every page after the first.

Name: _____

1. (Total: 12 points)

We fit a multiple regression model on the Advertising Data ¹ to predict sales (in thousands of units) for a particular product as a function of advertising budgets (in thousands of dollars) for TV, radio, and newspaper media. Below are the OLS results of the multiple linear regression of number of units sold on radio, TV, and newspaper advertising budgets.

OLS Regression Results						
=====						
Dep. Variable:	sales	R-squared:		0.915		
Model:	OLS	Adj. R-squared:		0.914		
Method:	Least Squares	F-statistic:		561.0		
Date:	Sat, 26 Jan 2019	Prob (F-statistic):		2.56e-83		
Time:	20:34:28	Log-Likelihood:		-299.34		
No. Observations:	160	AIC:		606.7		
Df Residuals:	156	BIC:		619.0		
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	3.0545	0.315	9.687	0.000	2.432	3.677
TV	0.0474	0.001	31.643	0.000	0.044	0.050
radio	0.1732	0.009	18.739	0.000	0.155	0.191
newspaper	0.0047	0.006	0.744	0.458	-0.008	0.017
=====						

¹The Advertising data set consists of the sales of that product in 200 different markets, along with advertising budgets for the product in each of those markets for three different media: TV, radio, and newspaper.

- (a) (2 points) Is there a relationship between advertising budget and sales?

Solution: Yes since $\text{Prob}(F\text{-statistic}) < 0.025$

- (b) (2 points) What is the average sales with \$120,000 marketing spent on TV advertising, \$30,000 on radio advertising, and \$10,000 on newspaper advertising?

Solution: Using the linear regression model, the avg sales is approximated as

$$\begin{aligned} & 3.0545 + 0.0475 * \text{Marketing Spend (in thousands of dollars) on TV} \\ & + 0.1732 * \text{Marketing Spend (in thousands of dollars) on radio} \\ & + 0.0047 * \text{Marketing Spend (in thousands of dollars) on newspaper} \end{aligned}$$

Thus,

$$\begin{aligned} \text{avg sales (in thousands of units)} & \approx 3.0545 + 0.0475 * 120 \\ & + 0.1732 * 30 \\ & + 0.0047 * 10 = 13.9855 \end{aligned}$$

and

$$\text{avg sales} \approx 13086$$

- (c) (2 points) How strong is the relationship between advertising budget and sales?

Solution: The R-squared is 0.915 showing a strong relationship between advertising budget and sales (91% of the variance in sales can be explained by the advertising budget).

- (d) (2 points) Which media contribute to sales? Do all three media contribute to sales, or do just one or two of the media contribute?

Solution: Only marketing on TV and Radio contribute to sales (since the p-values are less than 0.025)

- (e) (4 points) What is the average increase in sales for an additional \$1,000 spend in TV advertising? Provide a 95% confidence interval.

Solution: The coefficient for TV is estimated as 0.0475. Thus, an additional \$1,000 spent on TV advertising is associated with selling approximately 47.5 additional units of the product with 95% confidence interval $(47.5 - 2 * 1, 47.5 + 2 * 1) = (45.5, 49.5)$