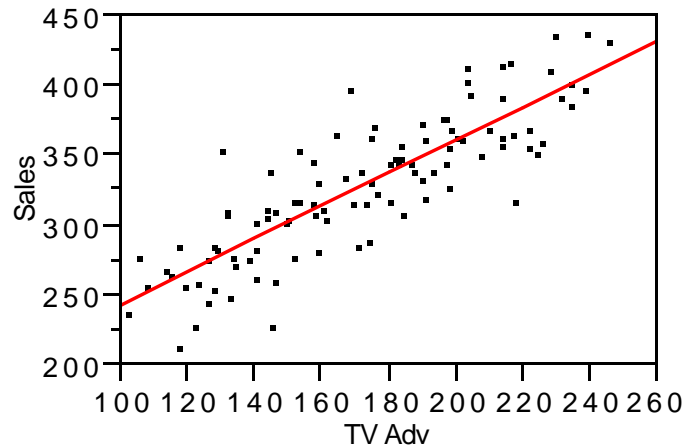


Collinearity in Regression Time Trends and Causation

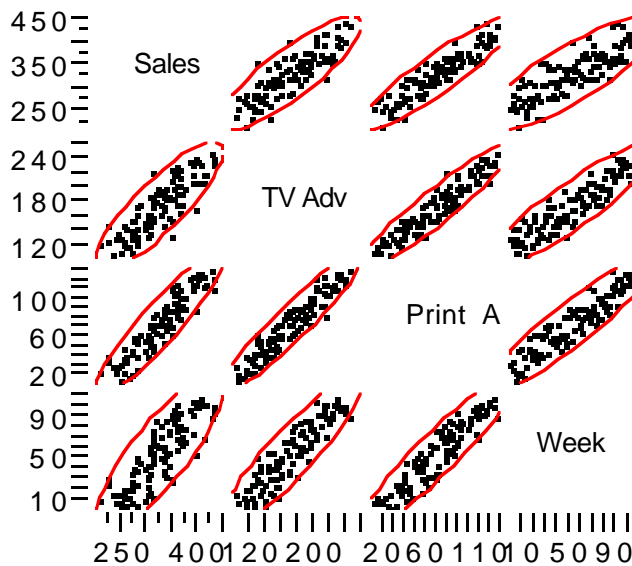
A rapidly growing firm would like to improve its allocation of advertising dollars between television and print media. Television now gets the largest share. Should this continue?

An initial marginal analysis quantifies the effect of television advertising. Indeed, the effect is strong ($R^2 = 0.72$) and positive. The slope suggests that each dollar spent on TV advertising earns on average an increase of \$1.18 in sales.



Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	124.09	12.71	9.76	<.0001
TV Adv	1.18	0.07	16.34	<.0001

The scatterplot matrix (with “time” in the last column to show time trends) indicates that both sales and TV spending have grown over the two years, but so has print advertising. The correlations between sales and these in the top row are 0.85, 0.91, and 0.83.



A multiple regression suggests a different impression for the effect of television advertising on sales. In contrast to the strong effect suggested in the simple regression on the previous page, TV advertising does *not* have a significant impact on sales in a multiple regression. Indeed, the estimated partial slope for TV advertising is slightly negative.

Summary of Fit

RSquare	0.84
Root Mean Square Error	20.71
Observations	104.00

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	228.91	16.04	14.27	<.0001
Print Adv	1.70	0.20	8.28	<.0001
TV Adv	-0.13	0.17	-0.78	0.4347

Are these differences “just statistical”, or do they also have substantive importance? The initial simple (or marginal) regression suggests that if the company were to increase TV advertising by, say, \$10 million, then its sales would be expected to grow by about \$11.8 million. The multiple regression reveals, however, that unless the firm also increased its print exposure in a manner commensurate with past practice, this increase would be for naught. Increased TV advertising – holding constant levels of print advertising – has no significant impact on sales (i.e., the partial slope for TV advertising is not significantly different from zero). Instead, the multiple regression shows that increased *print* advertising would have a strong effect even when TV advertising was left unchanged. The partial slope for printed advertising is positive and significant.

While this form of observed collinearity is troubling, it’s not the worst that can happen. What about the potential for having other sources of collinearity? Might there be other collinear factors hidden from our analysis? Sure. The only way we can begin to believe that we have the right model is to understand more of the context. The presence of so much collinearity among the predictors is a clue that surprising changes can occur in a regression when a new predictor, here printed advertising, is added to the model. Large amounts of collinearity are common when tracking a growing business or economy over time – everything is getting larger together.

Finally, we should *check assumptions*, particular for trends in the residuals that might suggest autocorrelation (dependence). We could then check for constant error variance (plot residuals on predicted), and lastly check for normality (a quantile plot of the residuals).