10 | Linear Regression, Part 3

Ivan Corneillet

Data Scientist

GENERAL ASSEMBLY

Learning Objectives

After this lesson, you should be able to:

- Explain what is one-hot encoding for categorical variables and how to use it for linear regression modeling
- Understand what are interaction effects and how to use the hierarchy principle for linear regression modeling

Linear Regression

One-Hot Encoding for Categorical Variables and SF Housing

Back to the SF housing dataset and the issue of beds and baths

- So far, we've considered *Beds* and *Baths* as ratio variables
  - Namely that the price premium between a property with 1 bathroom and another with 2 bathrooms was the same between a property with 3 bathrooms and another with 4 bathrooms
- Does this make sense?

Dep. Variable:	SalePrice	**R-squared:**	0.137
Model:	OLS	**Adj. R-squared:**	0.136
Method:	Least Squares	**F-statistic:**	146.6
Date:		**Prob (F-statistic):**	1.94e-31
Time:		**Log-Likelihood:**	-1690.7
No. Observations:	929	**AIC:**	3385.
Df Residuals:	927	**BIC:**	3395.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Intercept	0.3401	0.099	3.434	0.001	0.146 0.535
Baths	0.5242	0.043	12.109	0.000	0.439 0.609

Omnibus:	1692.623	**Durbin-Watson:**	1.582
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2167434.305
Skew:	12.317	**Prob(JB):**	0.00
Kurtosis:	238.345	**Cond. No.**	5.32

Back to the SF housing dataset and the issue of bed and bath counts (cont.)

- Let's test this hypothesis and convert $Baths$ to a nominal variable and then encode it into binary variables

m (# bathrooms)	$Bath = \begin{pmatrix} Bath_1, \\ Bath_2, \\ Bath_3, \\ Bath_4 \end{pmatrix}$ (one-hot encoding)
1	$(1, 0, 0, 0)$
2	$(0, 1, 0, 0)$
3	$(0, 0, 1, 0)$
4	$(0, 0, 0, 1)$

One-hot encoding for categorical variables

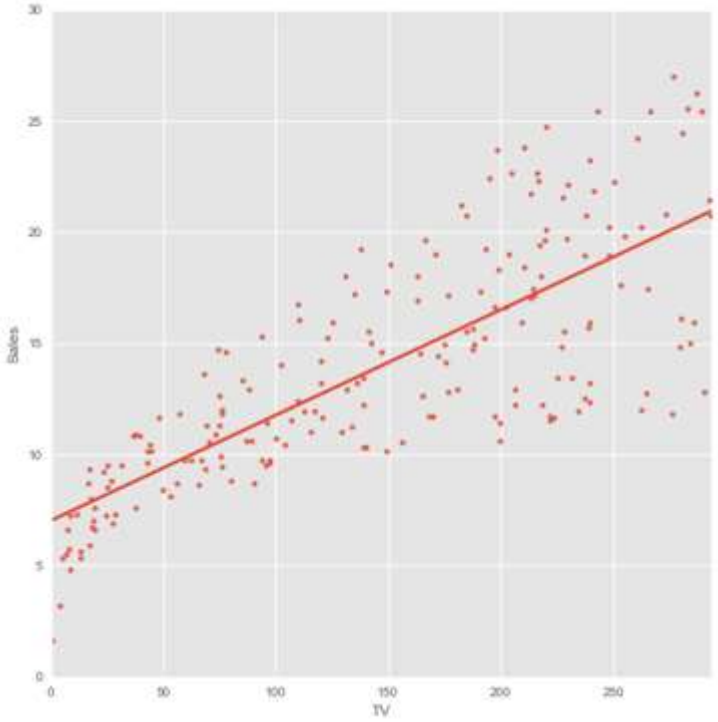
- This terminology from digital circuits where *one-hot* refers to a group of bits (here, our binary features) among which the legal combinations of values are only those with a single high (1) bit and all the others low (0)
- (Binary variables are also called *dummy* variables)

Linear Regression

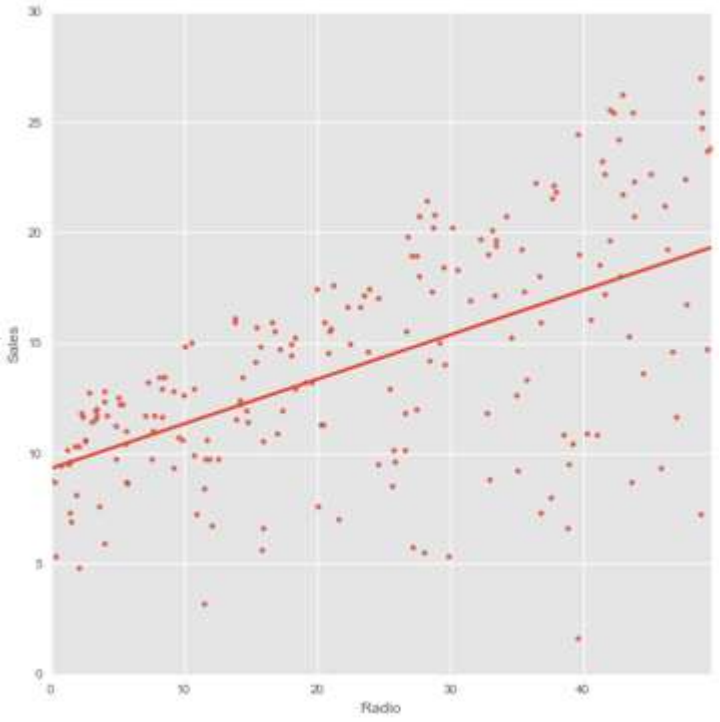
Interaction Effects and Advertising

Is there a relationship between advertising budget and sales?

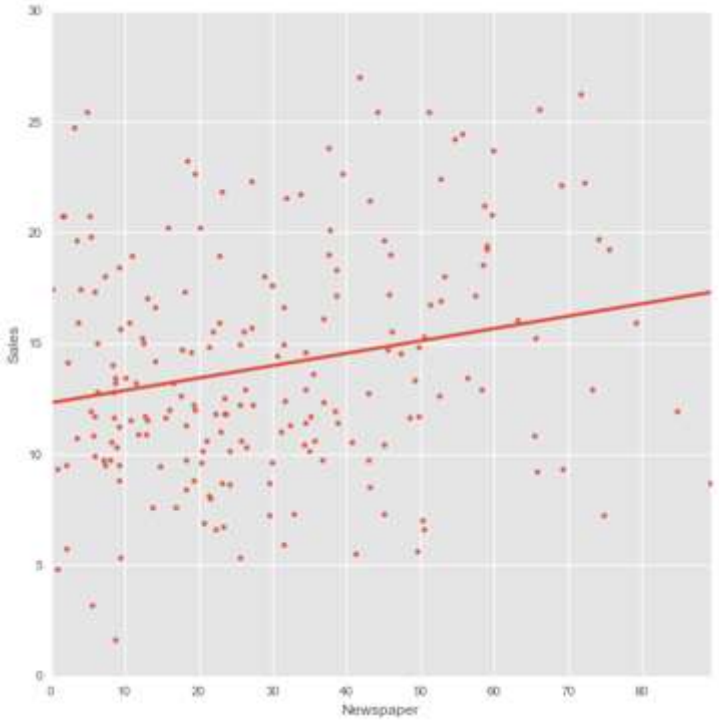
$Sales \sim TV$



$Sales \sim Radio$



$Sales \sim Newspaper$



Simple Linear Regressions on TV, $Radio$, and $Newspaper$

$Sales \sim TV$

Dep. Variable:	Sales	**R-squared:**	0.607
Model:	OLS	**Adj. R-squared:**	0.605
Method:	Least Squares	**F-statistic:**	302.8
Date:		**Prob (F-statistic):**	1.29e-41
Time:		**Log-Likelihood:**	-514.27
No. Observations:	198	**AIC:**	1033.
Df Residuals:	196	**BIC:**	1039.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	7.0306	0.462	15.219	0.000	6.120 7.942
TV	0.0474	0.003	17.400	0.000	0.042 0.053

Omnibus:	0.404	**Durbin-Watson:**	1.872
Prob(Omnibus):	0.817	**Jarque-Bera (JB):**	0.551
Skew:	-0.062	**Prob(JB):**	0.759
Kurtosis:	2.774	**Cond. No.**	338.

$Sales \sim Radio$

Dep. Variable:	Sales	**R-squared:**	0.333
Model:	OLS	**Adj. R-squared:**	0.329
Method:	Least Squares	**F-statistic:**	97.69
Date:		**Prob (F-statistic):**	5.99e-19
Time:		**Log-Likelihood:**	-566.70
No. Observations:	198	**AIC:**	1137.
Df Residuals:	196	**BIC:**	1144.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	9.3166	0.560	16.622	0.000	8.211 10.422
Radio	0.2016	0.020	9.884	0.000	0.161 0.242

Omnibus:	20.193	**Durbin-Watson:**	1.923
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	23.115
Skew:	-0.785	**Prob(JB):**	9.56e-06
Kurtosis:	3.582	**Cond. No.**	51.0

$Sales \sim Newspaper$

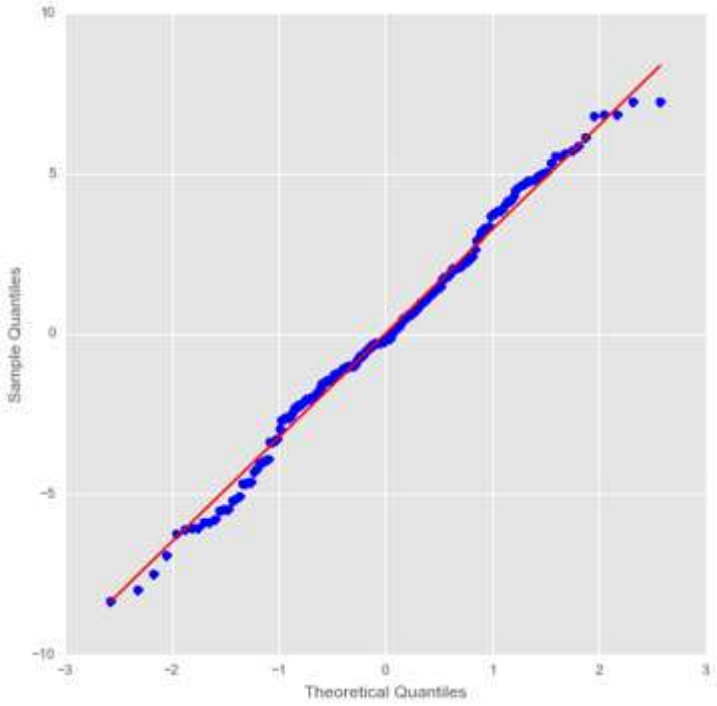
Dep. Variable:	Sales	**R-squared:**	0.048
Model:	OLS	**Adj. R-squared:**	0.043
Method:	Least Squares	**F-statistic:**	9.927
Date:		**Prob (F-statistic):**	0.00188
Time:		**Log-Likelihood:**	-601.84
No. Observations:	198	**AIC:**	1208.
Df Residuals:	196	**BIC:**	1214.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	12.3193	0.639	19.274	0.000	11.059 13.580
Newspaper	0.0558	0.018	3.151	0.002	0.021 0.091

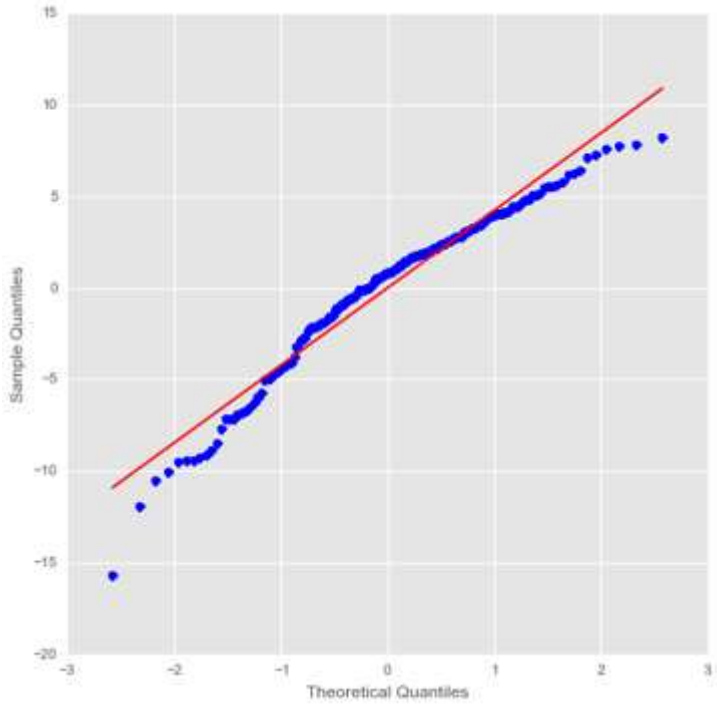
Omnibus:	5.835	**Durbin-Watson:**	1.916
Prob(Omnibus):	0.054	**Jarque-Bera (JB):**	5.303
Skew:	0.333	**Prob(JB):**	0.0706
Kurtosis:	2.555	**Cond. No.**	63.9

q-q plots of residuals. Are they normally distributed?

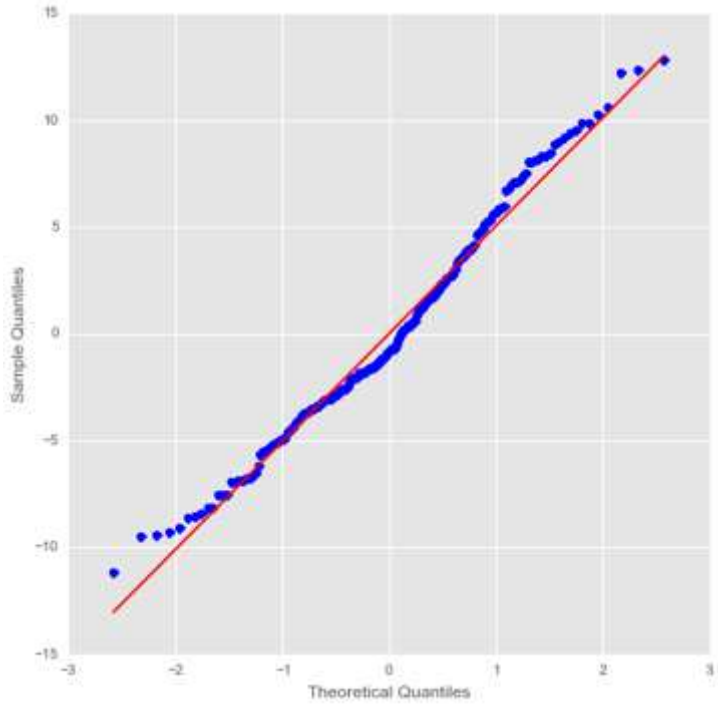
$Sales \sim TV$



$Sales \sim Radio$

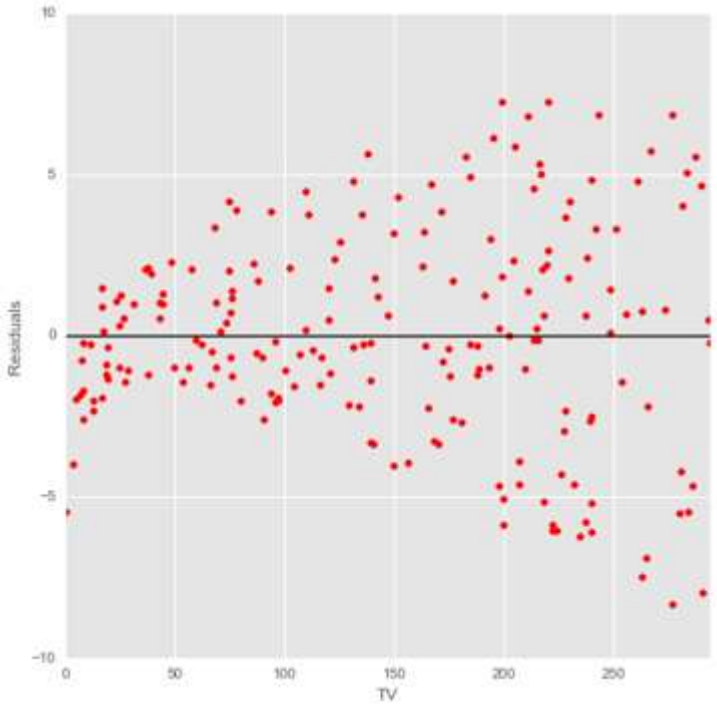


$Sales \sim Newspaper$

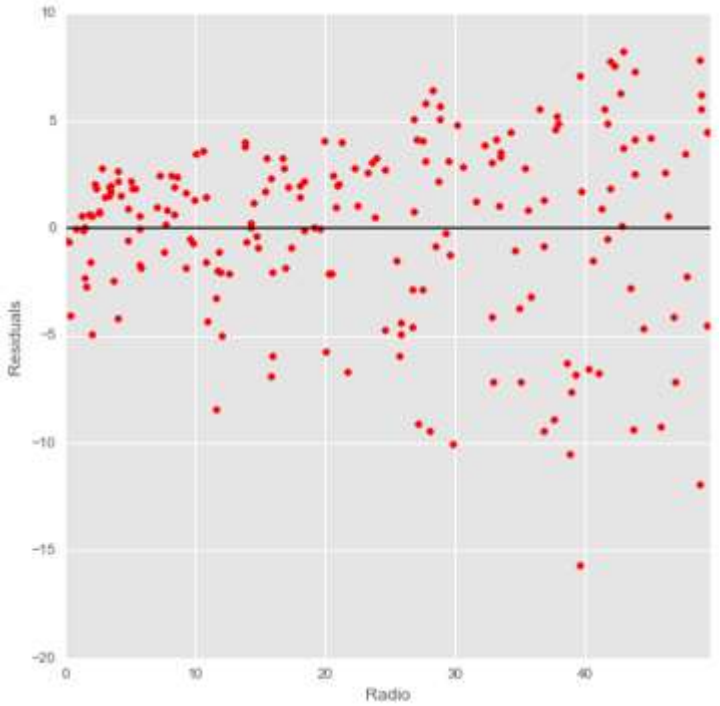


Scatterplots of residuals against advertising budget. Are they randomly distributed?

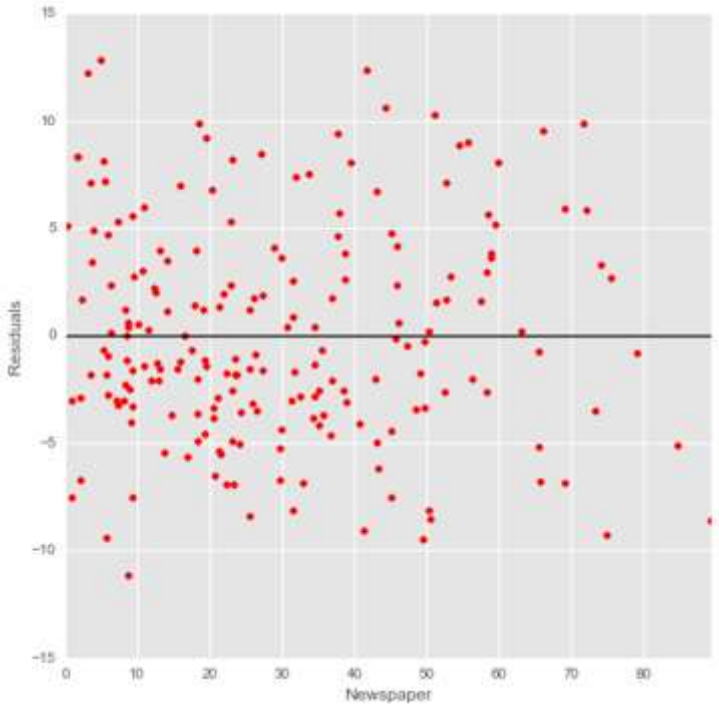
$Sales \sim TV$



$Sales \sim Radio$



$Sales \sim Newspaper$



$Sales \sim TV + Radio + Newspaper$

Dep. Variable:	Sales	**R-squared:**	0.895
Model:	OLS	**Adj. R-squared:**	0.894
Method:	Least Squares	**F-statistic:**	553.5
Date:		**Prob (F-statistic):**	8.35e-95
Time:		**Log-Likelihood:**	-383.24
No. Observations:	198	**AIC:**	774.5
Df Residuals:	194	**BIC:**	787.6
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	2.9523	0.318	9.280	0.000	2.325 3.580
TV	0.0457	0.001	32.293	0.000	0.043 0.048
Radio	0.1886	0.009	21.772	0.000	0.171 0.206
Newspaper	-0.0012	0.006	-0.187	0.852	-0.014 0.011

Omnibus:	59.593	**Durbin-Watson:**	2.041
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	147.654
Skew:	-1.324	**Prob(JB):**	8.66e-33
Kurtosis:	6.299	**Cond. No.**	457.

$Sales \sim TV + Radio$. Are we done yet?

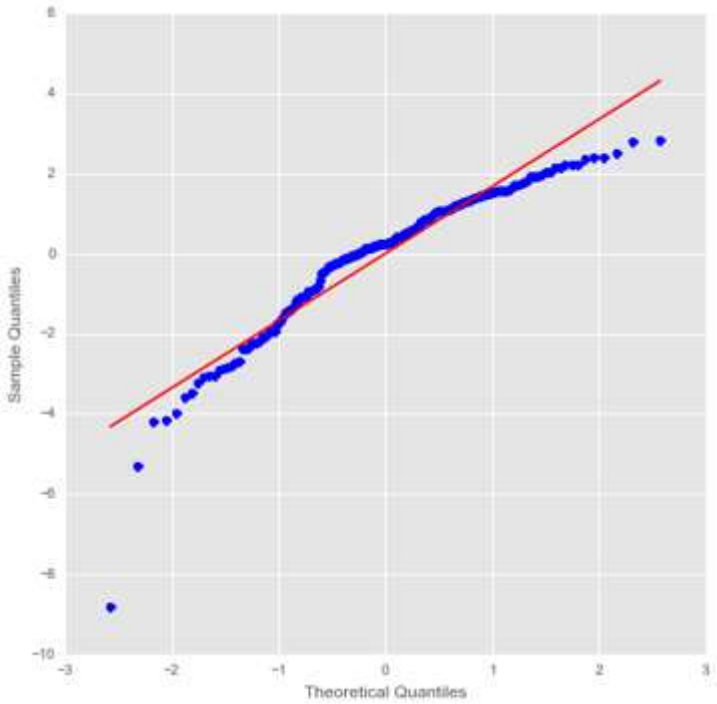
Dep. Variable:	Sales	**R-squared:**	0.895
Model:	OLS	**Adj. R-squared:**	0.894
Method:	Least Squares	**F-statistic:**	834.4
Date:		**Prob (F-statistic):**	2.60e-96
Time:		**Log-Likelihood:**	-383.26
No. Observations:	198	**AIC:**	772.5
Df Residuals:	195	**BIC:**	782.4
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	2.9315	0.297	9.861	0.000	2.345 3.518
TV	0.0457	0.001	32.385	0.000	0.043 0.048
Radio	0.1880	0.008	23.182	0.000	0.172 0.204

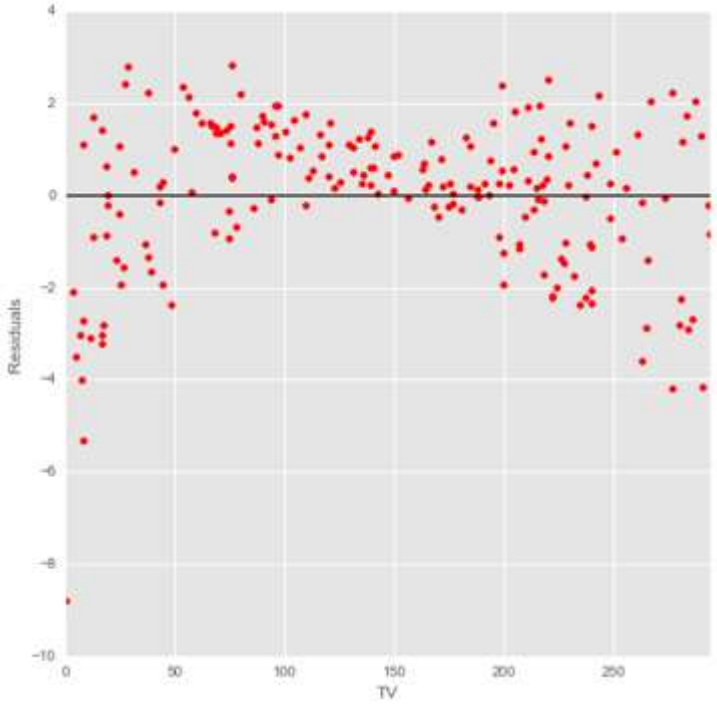
Omnibus:	59.228	**Durbin-Watson:**	2.038
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	145.127
Skew:	-1.321	**Prob(JB):**	3.06e-32
Kurtosis:	6.257	**Cond. No.**	423.

$Sales \sim TV + Radio$. What do you observe? Are we done yet?

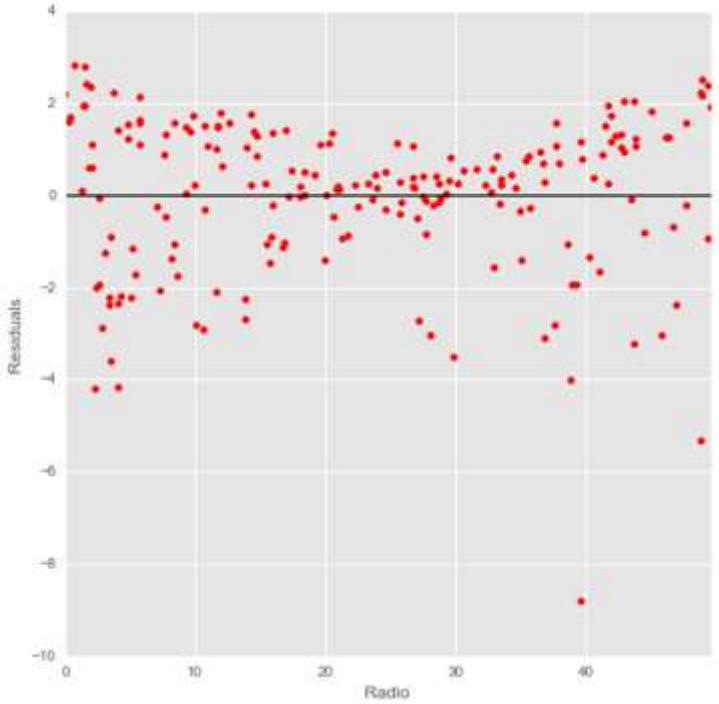
Residuals q-q plot



Residuals against TV



Residuals against $Radio$



$Sales \sim TV + Radio$

- $Sales = \underbrace{2.93}_{\widehat{\beta}_0} + \underbrace{.0457}_{\widehat{\beta}_1} \times TV + \underbrace{.188}_{\widehat{\beta}_2} \times Radio$

- This model assumes that the effect on sales of increasing one media (e.g., TV) is independent of the amount spent on the other media (e.g., $Radio$)

- More specifically, the model states that the average effect on sales of a one-unit increase ($\$1k$) in TV is always ($\underbrace{.0457}_{\widehat{\beta}_1}$ [$1k$ units] = 46 units), regardless of the amount spend on $Radio$

Interaction Effects

- But suppose that spending money on radio advertising actually increases the effectiveness of TV advertising
  - the slope term for TV should increase as $Radio$ increases
- E.g., given a fixed budget of $\$100k$, spending half on TV and half on radio may increase sales more than allocating the entire amount to either TV or radio
- This is known as a synergy effect in marketing; in statistics it is referred to as an interaction effect

Sales ~ TV + Radio + TV * Radio

Dep. Variable:	Sales	**R-squared:**	0.968
Model:	OLS	**Adj. R-squared:**	0.967
Method:	Least Squares	**F-statistic:**	1934.
Date:		**Prob (F-statistic):**	3.19e-144
Time:		**Log-Likelihood:**	-267.07
No. Observations:	198	**AIC:**	542.1
Df Residuals:	194	**BIC:**	555.3
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	6.7577	0.247	27.304	0.000	6.270 7.246
TV	0.0190	0.002	12.682	0.000	0.016 0.022
Radio	0.0276	0.009	3.089	0.002	0.010 0.045
TV:Radio	0.0011	5.27e-05	20.817	0.000	0.001 0.001

Omnibus:	126.182	**Durbin-Watson:**	2.241
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1151.060
Skew:	-2.306	**Prob(JB):**	1.12e-250
Kurtosis:	13.875	**Cond. No.**	1.78e+04

Interaction Effects (cont.)

- $Sales = \underbrace{6.76}_{\widehat{\beta}_0'} + \underbrace{.0190}_{\widehat{\beta}_1'} \times TV + \underbrace{.0276}_{\widehat{\beta}_2'} \times Radio + \underbrace{.0011}_{\widehat{\beta}_3'} \times TV \times Radio$
- The interaction is important
  - β_3' is statistically significant
  - R^2 with this model went up to 96.8% up from 89.5% for the model without interaction. This that $1 - \frac{1-.968}{1-.895} = .70 = 70\%$ of the unexplained variability in the previous model has been explained by the interaction term

Hierarchy Principle

- Sometimes an interaction term $x_i \cdot x_j$ is significant, but one or both of its main effects (in this case x_i and/or x_j) are not

- The hierarchy principle
  - If we include an interaction in a model, we should also include the main effects, even if they aren't significant