Linear Regression, Part 3

Ivan Corneillet

Data Scientist

GENERAL ASSEMBLY

Today, we will go deeper in a few linear regression topics, each with a different dataset:

- Model Fit and Customer Retention
- One-Hot Encoding for Categorical Variables and SF Housing
- Interaction Effects and Advertising
  - Hierarchy Principle

Review

Underfit, optimal fit, and overfit

Activity | Underfit, optimal fit, and overfit

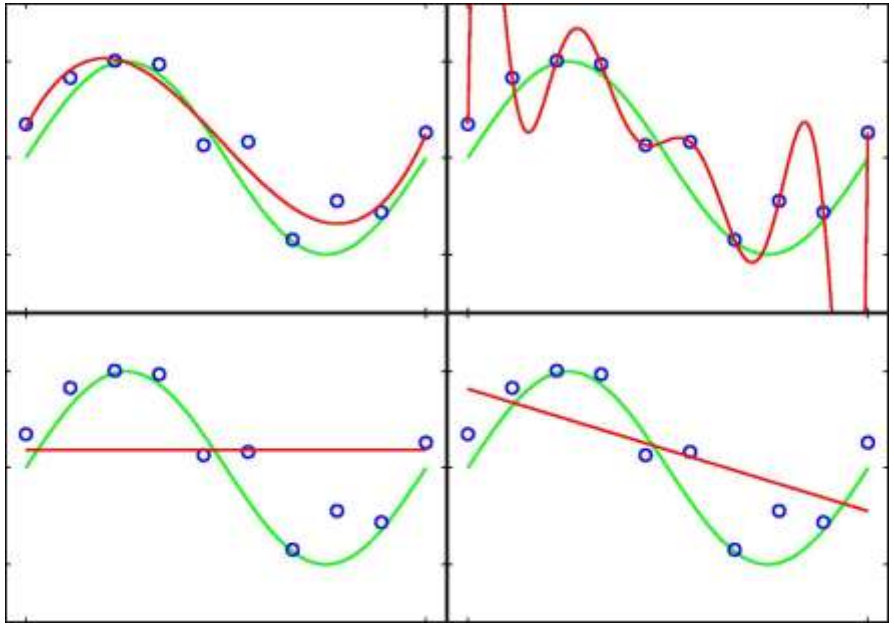
EXERCISE

DIRECTIONS (5 minutes)

1. Classify the following polynomial regressions according to their fit:
    1. Underfit
    2. Optimal fit
    3. Overfit
2. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions



Linear Regression

Model Fit and Customer Retention

Activity | Customer Retention

EXERCISE

DIRECTIONS (20 minutes)

1. The following dataset documents the “survival” pattern over seven years for a sample of 1000 customers who were all “acquired” in the same period
2. Build one or more models to capture this pattern, then use each model to project the survival curve over the next five years
3. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

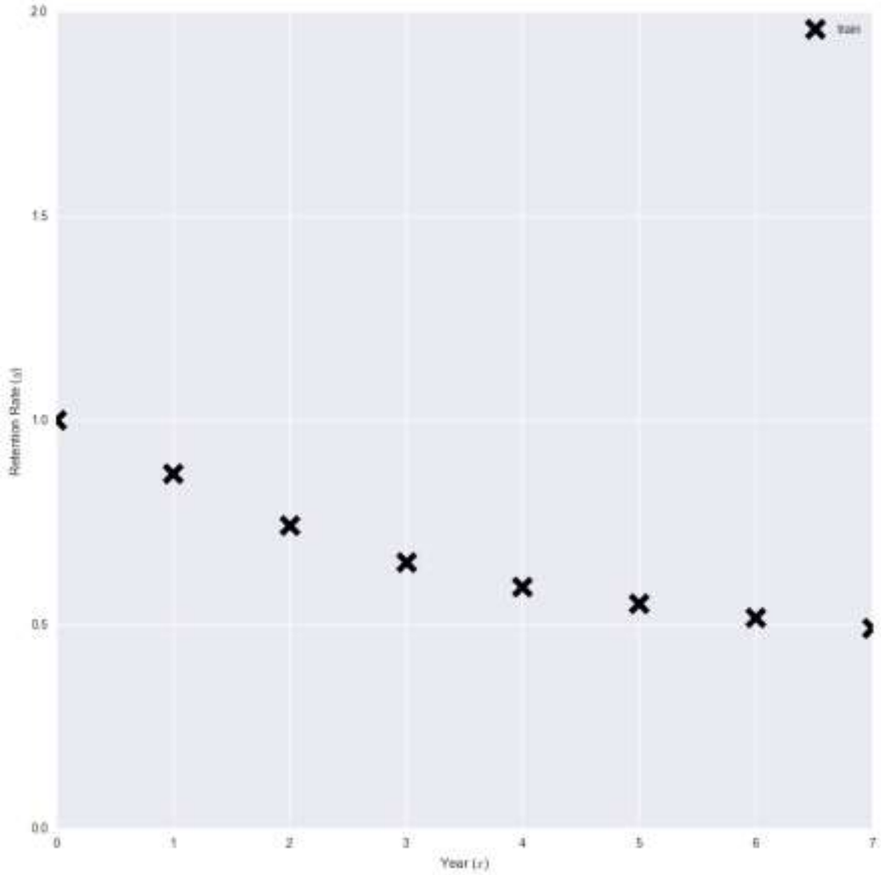
Year	Retention Rate
0	1
1	.869
2	.743
3	.653
4	.593
5	.551
6	.517
7	.491

Source: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management

Activity | Retention rate (y) as a function of the year (x)

EXERCISE

Year (x)	Retention Rate (y)
0	1
1	.869
2	.743
3	.653
4	.593
5	.551
6	.517
7	.491



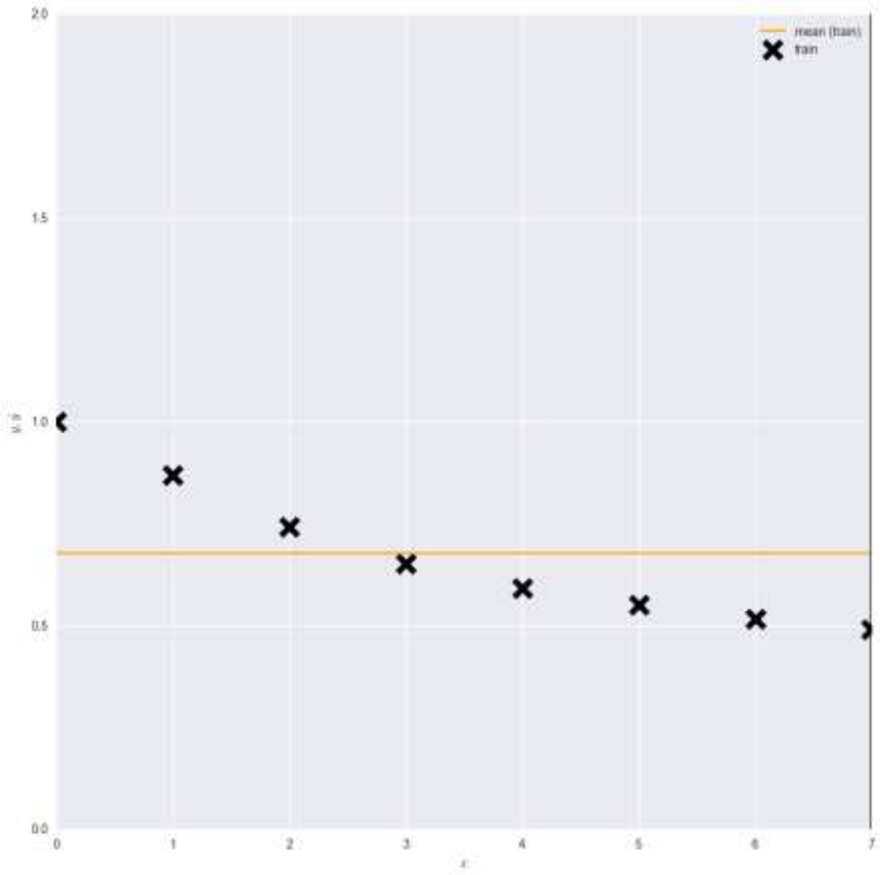
Activity | Linear Model: $y = .6771$

EXERCISE

Dep. Variable:	y	R-squared:	0.000
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	nan
Date:		Prob (F-statistic):	nan
Time:		Log-Likelihood:	2.8580
No. Observations:	8	AIC:	-3.716
Df Residuals:	7	BIC:	-3.637
Df Model:	0		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	0.6771	0.064	10.583	0.000	0.526 0.828

Omnibus:	1.441	Durbin-Watson:	0.211
Prob(Omnibus):	0.486	Jarque-Bera (JB):	0.904
Skew:	0.717	Prob(JB):	0.636
Kurtosis:	2.192	Cond. No.	1.00



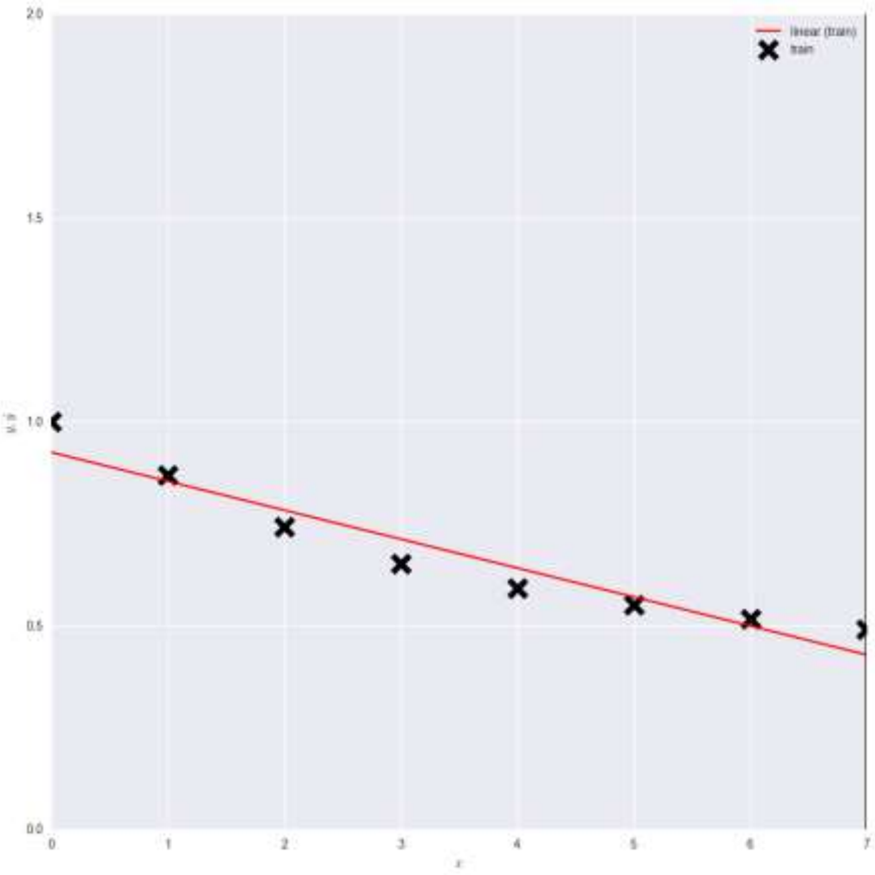
Activity | Linear Model: $y = .9254 - .0709 \cdot t$

EXERCISE

Dep. Variable:	y	**R-squared:**	0.922
Model:	OLS	**Adj. R-squared:**	0.909
Method:	Least Squares	**F-statistic:**	70.91
Date:		**Prob (F-statistic):**	0.000153
Time:		**Log-Likelihood:**	13.061
No. Observations:	8	**AIC:**	-22.12
Df Residuals:	6	**BIC:**	-21.96
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	0.9254	0.035	26.258	0.000	0.839 1.012
x	-0.0709	0.008	-8.421	0.000	-0.092 -0.050

Omnibus:	1.277	**Durbin-Watson:**	0.634
Prob(Omnibus):	0.528	**Jarque-Bera (JB):**	0.711
Skew:	0.310	**Prob(JB):**	0.701
Kurtosis:	1.678	**Cond. No.**	7.95



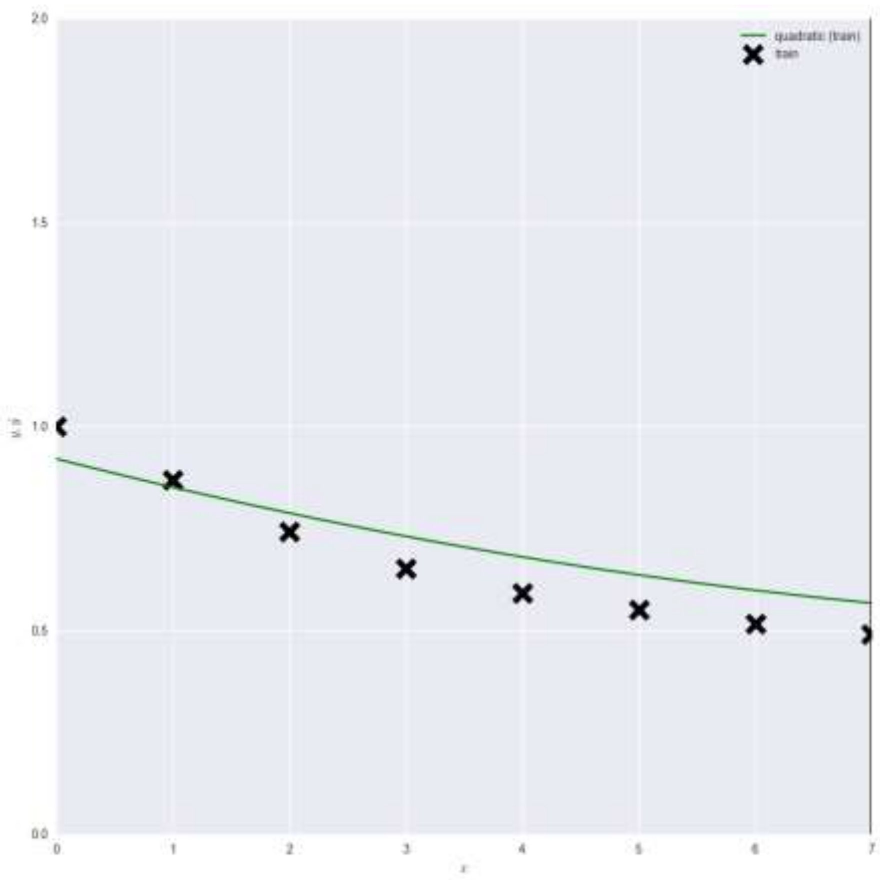
Activity | Quadratic Model: $y = .9211 - .0729 \cdot t + .0032 \cdot t^2$

EXERCISE

Dep. Variable:	y	**R-squared:**	0.923
Model:	OLS	**Adj. R-squared:**	0.892
Method:	Least Squares	**F-statistic:**	30.03
Date:		**Prob (F-statistic):**	0.00164
Time:		**Log-Likelihood:**	13.121
No. Observations:	8	**AIC:**	-20.24
Df Residuals:	5	**BIC:**	-20.00
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	0.9211	0.041	22.274	0.000	0.815 1.027
x	-0.0729	0.012	-6.252	0.002	-0.103 -0.043
x ^ 2	0.0032	0.012	0.275	0.795	-0.027 0.033

Omnibus:	1.491	**Durbin-Watson:**	0.630
Prob(Omnibus):	0.474	**Jarque-Bera (JB):**	0.769
Skew:	0.342	**Prob(JB):**	0.681
Kurtosis:	1.644	**Cond. No.**	11.6



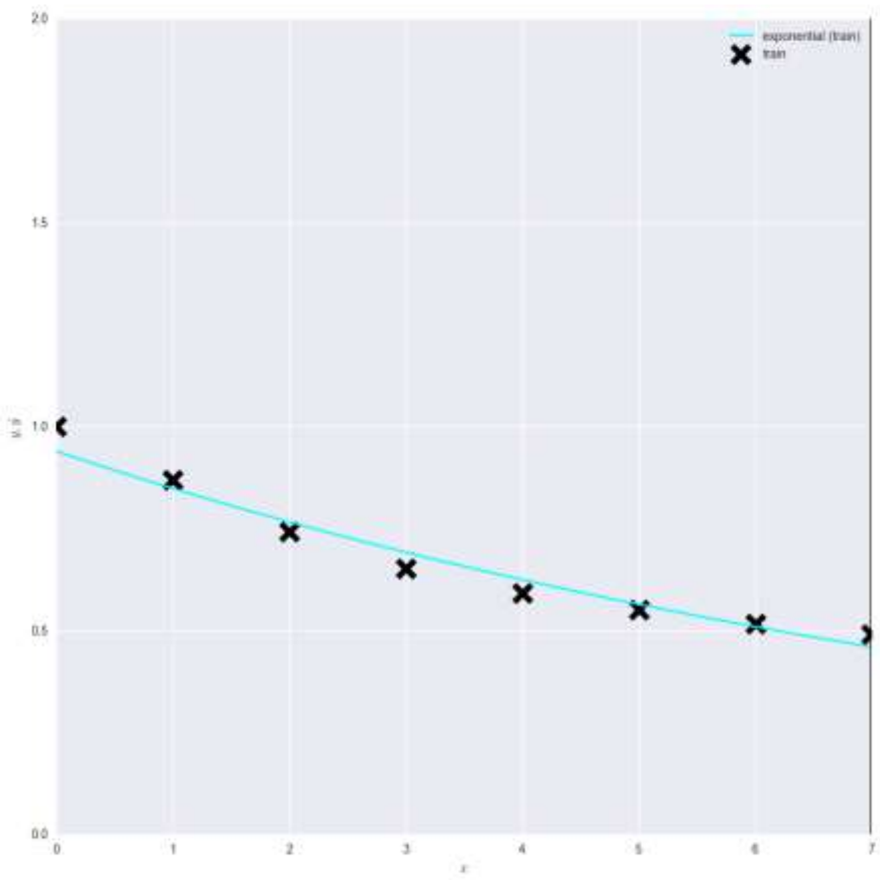
Activity | Exponential Model: $ln(y) = -.0621 - .1020 \cdot t$

EXERCISE

Dep. Variable:	log_y	**R-squared:**	0.964
Model:	OLS	**Adj. R-squared:**	0.958
Method:	Least Squares	**F-statistic:**	159.1
Date:		**Prob (F-statistic):**	1.52e-05
Time:		**Log-Likelihood:**	13.389
No. Observations:	8	**AIC:**	-22.78
Df Residuals:	6	**BIC:**	-22.62
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Intercept	-0.0621	0.034	-1.836	0.116	-0.145 0.021
x	-0.1020	0.008	-12.615	0.000	-0.122 -0.082

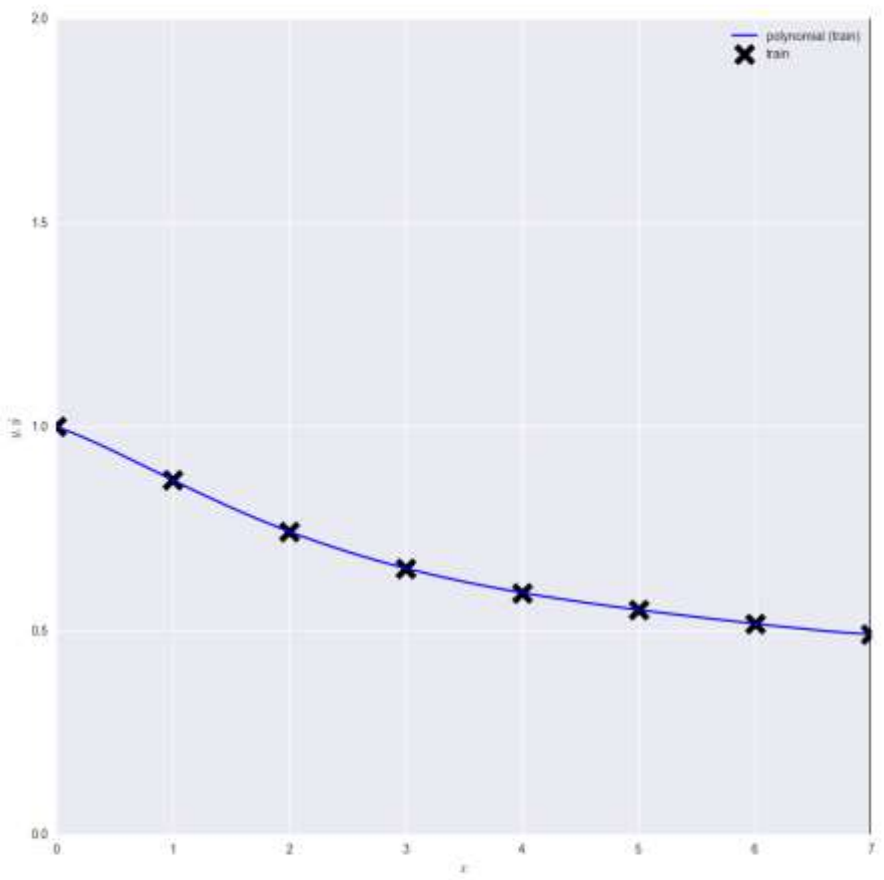
Omnibus:	1.695	**Durbin-Watson:**	0.610
Prob(Omnibus):	0.429	**Jarque-Bera (JB):**	0.739
Skew:	0.196	**Prob(JB):**	0.691
Kurtosis:	1.564	**Cond. No.**	7.95



Activity | Polynomial of degree 7

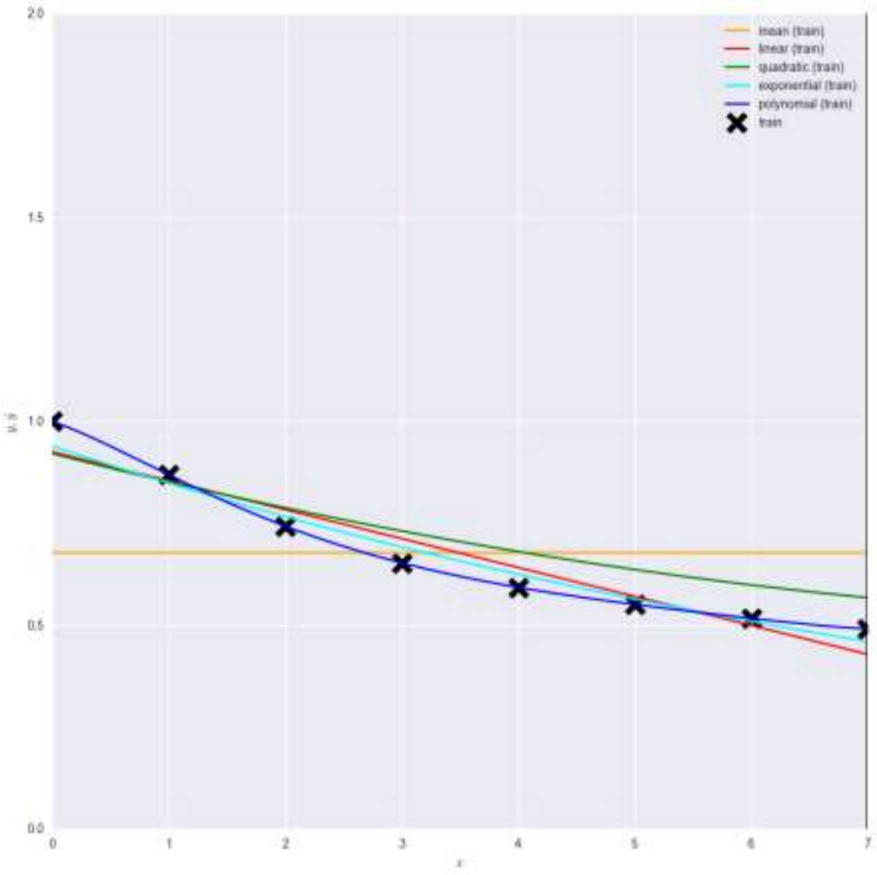
EXERCISE

$$
\begin{gathered}
y = \\
1 \\
-.100597619 \cdot t \\
-.0596777778 \cdot t^2 \\
+.0380569444 \cdot t^3 \\
-.0101944444 \cdot t^4 \\
+.00153611111 \cdot t^5 \\
-.0001277777 \cdot t^6 \\
+.00000456349206 \cdot t^7
\end{gathered}
$$



Activity | Training data's score (R^2) (using *sklearn*)

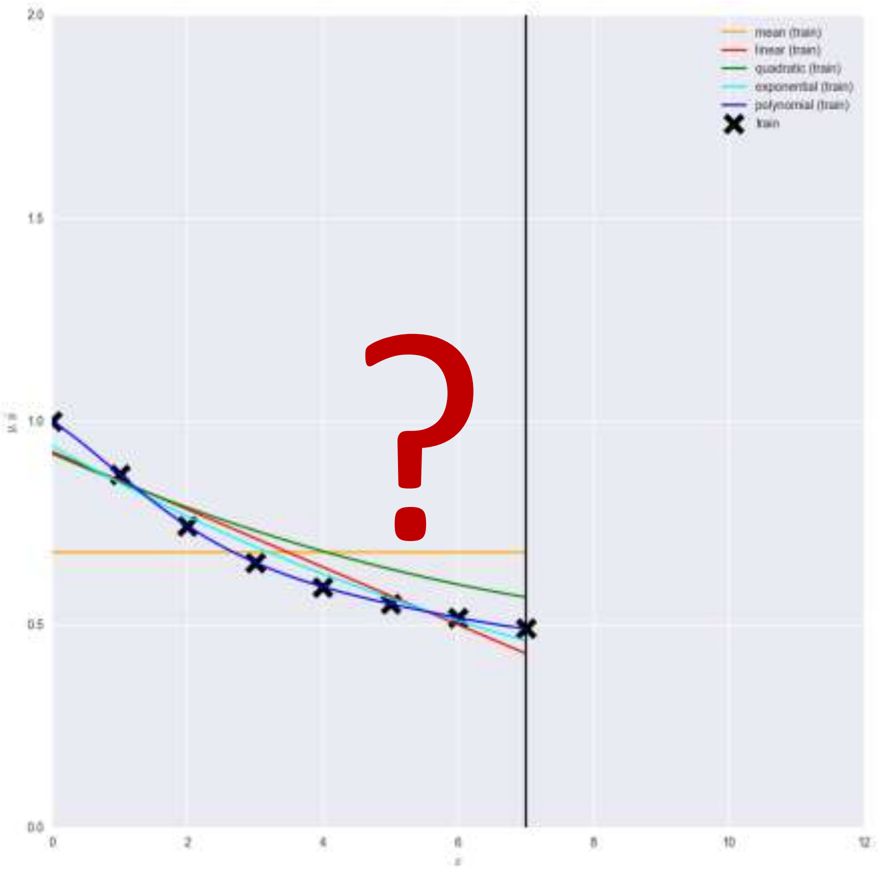
EXERCISE



Model	Training score (R^2)
Mean	0
Linear	.999
Quadratic	.923
Exponential	.964
Polynomial	1

Activity | Test data and what happens to R^2?

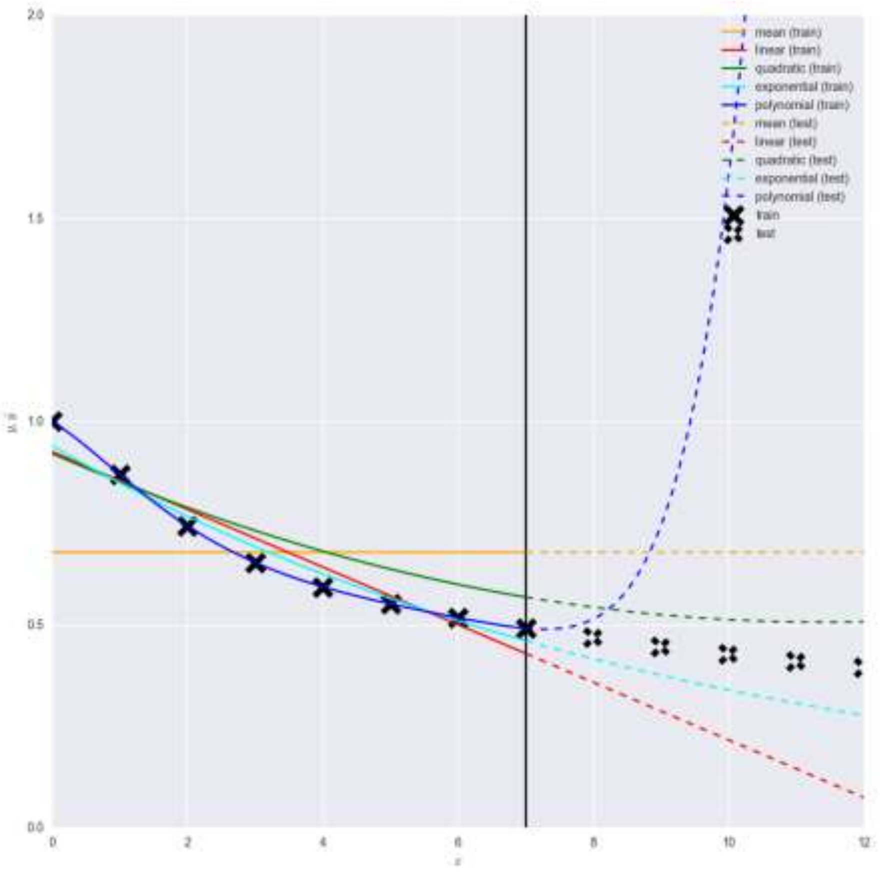
EXERCISE



Model	Testing score (R^2)
Mean	
Linear	
Quadratic	?
Exponential	
Polynomial	

Activity | Test data and what happens to R^2? (cont.) (using *sklearn*)

EXERCISE



Model	Training score ($\boldsymbol{R^2}$)
Mean	-90
Linear	-73
Quadratic	-68
Exponential	-11
Polynomial	-30140

Linear Regression

One-Hot Encoding for Categorical Variables and SF Housing

Back to the SF housing dataset and the issue of beds and baths (TODO)

- 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 beds and baths (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	
2	
3	
4	

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	
3	
4	

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	
4	

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	

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)

Activity | One-hot encoding for the $Baths$ categorical feature

EXERCISE

DIRECTIONS (10 minutes)

1. Run the 4 regressions highlighted in the codealong (Part A). Each regression includes only 3 out of the 4 binary variables we created
2. How do you interpret the βs in each regression?
3. Why do we only need 3 binary variables, not all 4?
4. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

Activity | One-hot encoding for the $Baths$ categorical feature (cont.)

EXERCISE

$$SalePrice = \beta_1 + \beta_{1,2} \cdot Bath_2 + \beta_{1,3} \cdot Bath_3 + \beta_{1,4} \cdot Bath_4$$

(don't include $Bath_1$)

$$SalePrice = \beta_2 + \beta_{2,1} \cdot Bath_1 + \beta_{2,3} \cdot Bath_3 + \beta_{2,4} \cdot Bath_4$$

(don't include $Bath_2$)

$$SalePrice = \beta_3 + \beta_{3,1} \cdot Bath_1 + \beta_{3,2} \cdot Bath_2 + \beta_{3,4} \cdot Bath_4$$

(don't include $Bath_3$)

$$SalePrice = \beta_4 + \beta_{4,1} \cdot Bath_1 + \beta_{4,2} \cdot Bath_2 + \beta_{4,3} \cdot Bath_3$$

(don't include $Bath_4$)

One-hot encoding for the $Baths$ categorical feature (cont.)

Dep. Variable:	SalePrice	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.039
Method:	Least Squares	**F-statistic:**	11.78
Date:		**Prob (F-statistic):**	1.49e-07
Time:		**Log-Likelihood:**	-1314.2
No. Observations:	794	**AIC:**	2636.
Df Residuals:	790	**BIC:**	2655.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	0.9914	0.070	14.249	0.000	0.855 1.128
Bath_2	0.2831	0.099	2.855	0.004	0.088 0.478
Bath_3	0.4808	0.142	3.383	0.001	0.202 0.760
Bath_4	1.2120	0.232	5.231	0.000	0.757 1.667

Omnibus:	1817.972	**Durbin-Watson:**	1.867
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8069883.811
Skew:	19.917	**Prob(JB):**	0.00
Kurtosis:	495.280	**Cond. No.**	5.79

Dep. Variable:	SalePrice	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.039
Method:	Least Squares	**F-statistic:**	11.78
Date:		**Prob (F-statistic):**	1.49e-07
Time:		**Log-Likelihood:**	-1314.2
No. Observations:	794	**AIC:**	2636.
Df Residuals:	790	**BIC:**	2655.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	1.2745	0.071	18.040	0.000	1.136 1.413
Bath_1	-0.2831	0.099	-2.855	0.004	-0.478 -0.088
Bath_3	0.1977	0.143	1.386	0.166	-0.082 0.478
Bath_4	0.9290	0.232	4.003	0.000	0.473 1.384

Omnibus:	1817.972	**Durbin-Watson:**	1.867
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8069883.811
Skew:	19.917	**Prob(JB):**	0.00
Kurtosis:	495.280	**Cond. No.**	5.84

Dep. Variable:	SalePrice	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.039
Method:	Least Squares	**F-statistic:**	11.78
Date:		**Prob (F-statistic):**	1.49e-07
Time:		**Log-Likelihood:**	-1314.2
No. Observations:	794	**AIC:**	2636.
Df Residuals:	790	**BIC:**	2655.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	1.4722	0.124	11.881	0.000	1.229 1.715
Bath_1	-0.4808	0.142	-3.383	0.001	-0.760 -0.202
Bath_2	-0.1977	0.143	-1.386	0.166	-0.478 0.082
Bath_4	0.7313	0.253	2.886	0.004	0.234 1.229

Omnibus:	1817.972	**Durbin-Watson:**	1.867
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8069883.811
Skew:	19.917	**Prob(JB):**	0.00
Kurtosis:	495.280	**Cond. No.**	7.52

Dep. Variable:	SalePrice	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.039
Method:	Least Squares	**F-statistic:**	11.78
Date:		**Prob (F-statistic):**	1.49e-07
Time:		**Log-Likelihood:**	-1314.2
No. Observations:	794	**AIC:**	2636.
Df Residuals:	790	**BIC:**	2655.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	2.2035	0.221	9.969	0.000	1.770 2.637
Bath_1	-1.2120	0.232	-5.231	0.000	-1.667 -0.757
Bath_2	-0.9290	0.232	-4.003	0.000	-1.384 -0.473
Bath_3	-0.7313	0.253	-2.886	0.004	-1.229 -0.234

Omnibus:	1817.972	**Durbin-Watson:**	1.867
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8069883.811
Skew:	19.917	**Prob(JB):**	0.00
Kurtosis:	495.280	**Cond. No.**	11.7

Activity | One-hot encoding for the $Baths$ categorical feature (cont.)

EXERCISE

β_1		$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$
β_2	$\beta_{2,1}$		$\beta_{2,3}$	$\beta_{2,4}$
β_3	$\beta_{3,1}$	$\beta_{3,2}$		$\beta_{3,4}$
β_4	$\beta_{4,1}$	$\beta_{4,2}$	$\beta_{4,3}$	

Activity | One-hot encoding for the $Baths$ categorical feature (cont.)

EXERCISE

β_1 0.9914		$\beta_{1,2}$ 0.2831	$\beta_{1,3}$ 0.4808	$\beta_{1,4}$ 1.212
β_2 1.2745	$\beta_{2,1}$ –0.2831		$\beta_{2,3}$ 0.1977	$\beta_{2,4}$ 0.9290
β_3 1.4722	$\beta_{3,1}$ –0.4808	$\beta_{3,2}$ –0.1977		$\beta_{3,4}$ 0.7313
β_4 2.2025	$\beta_{4,1}$ –1.212	$\beta_{4,2}$ –0.9290	$\beta_{4,3}$ –0.7313	

Activity | One-hot encoding for the $Baths$ categorical feature (cont.)

EXERCISE

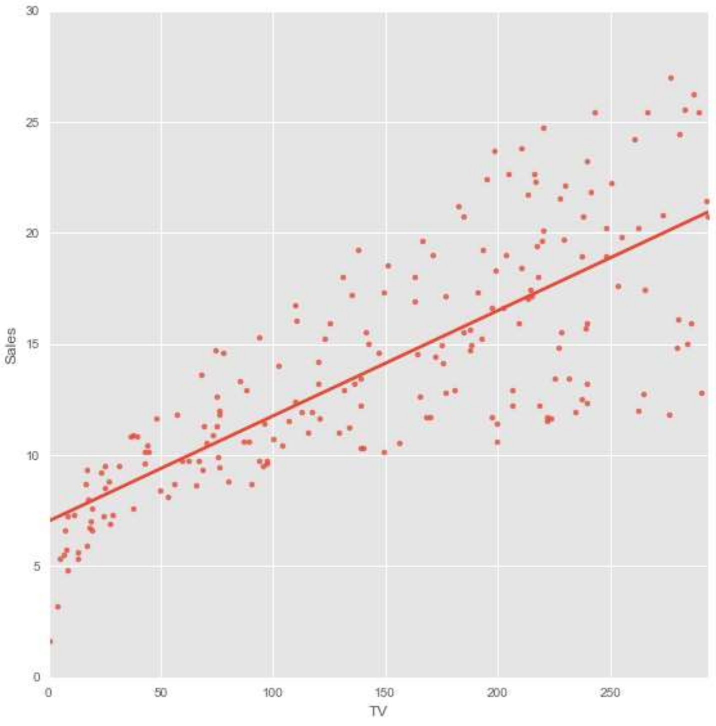
β_i	Value (Sale's price) of a property in SF with i bathrooms
$\beta_{i,j}$ when $j > i$	Increase of value for a property when increasing the number of bathrooms from i to j (while keeping the rest of the same…)
$\beta_{i,j}$ when $j < i$	Decrease of value for a property when decreasing the number of bathrooms from i to j (while keeping the rest of the same…)
$\beta_{i,j} = -\beta_{j,i}$	Going from i to j bathrooms has the opposite effect of going from j bathrooms to i bathrooms
$\beta_j = \beta_i + \beta_{i,j}$ for any i and j	E.g., $\beta_4 = \beta_1 + \beta_{1,4}$. I.e., the value of a 4 bathrooms can be derived from a 1 bedroom house and by increasing the number of bathrooms for 1 to 4
$\beta_{i,j} = \beta_{i,k} + \beta_{k,j}$ for any i, j and k	E.g., $\beta_{1,4} = \beta_{1,2} + \beta_{2,4}$. I.e., the increase in value from a 1 bathroom house to a 4 bathrooms house is identical to going from upgrading from 1 bathroom to 2 bathrooms and then from upgrading from 2 bathrooms to 4 bathrooms

Linear Regression

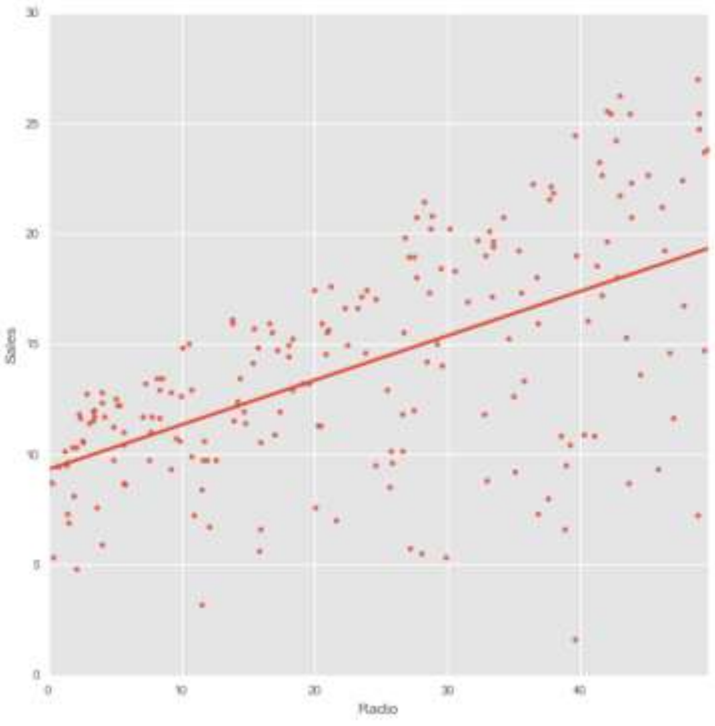
Interaction Effects and Advertising

Is there a relationship between advertising budget and sales?

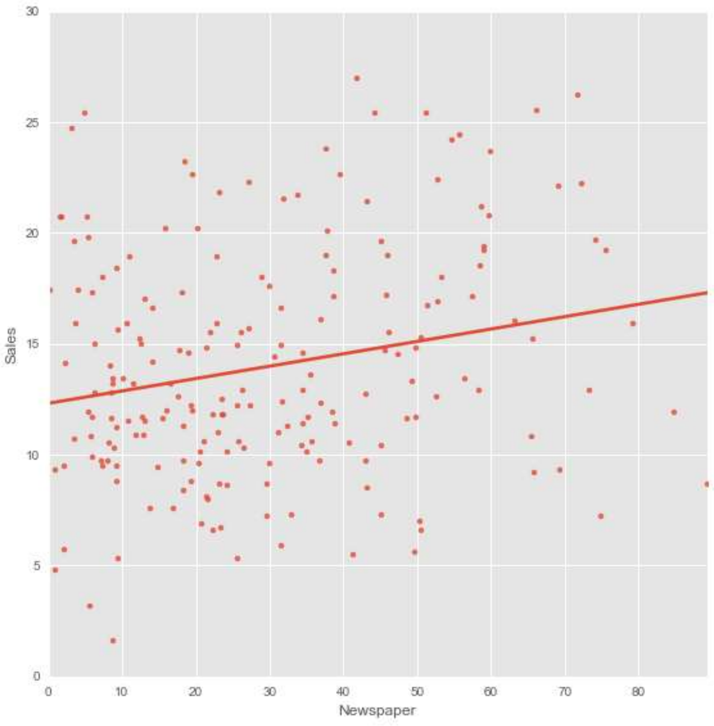
$Sales \sim TV$



$Sales \sim Radio$



$Sales \sim Newspaper$



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

$Sales \sim Newspaper$

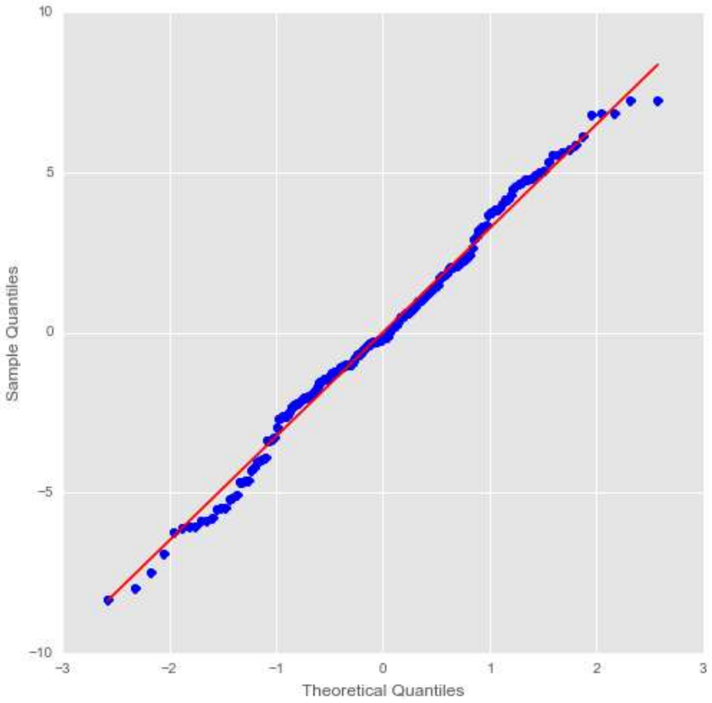
Dep. Variable:	Sales	R-squared:	0.048
Model:	OLS	Adj. R-squared:	0.043
Method:	Least Squares	F-statistic:	9.927
Date:		Prob (F-statistic):	0.00188
Time:		Log-Likelihood:	-601.84
No. Observations:	198	AIC:	1208.
Df Residuals:	196	BIC:	1214.
Df Model:	1		
Covariance Type:	nonrobust		

	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

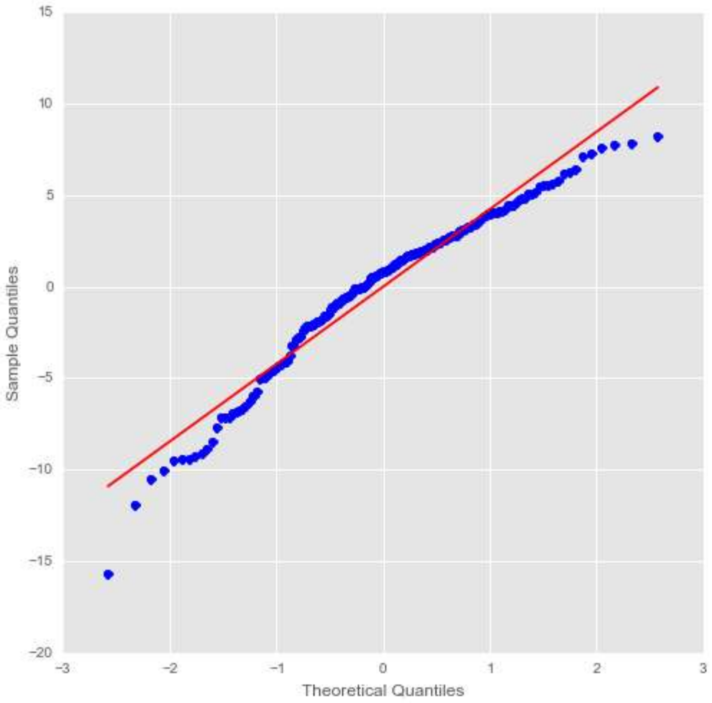
Omnibus:	5.835	Durbin-Watson:	1.916
Prob(Omnibus):	0.054	Jarque-Bera (JB):	5.303
Skew:	0.333	Prob(JB):	0.0706
Kurtosis:	2.555	Cond. No.	63.9

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

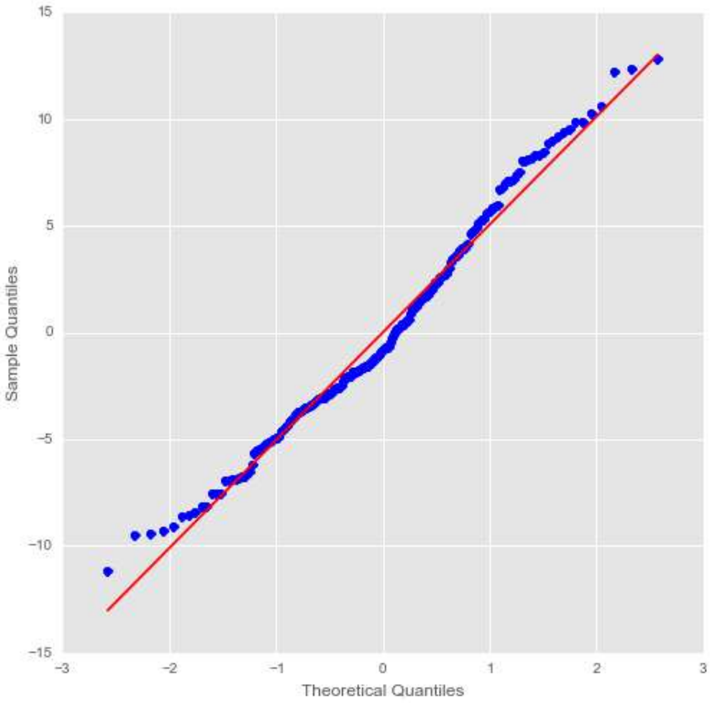
$Sales \sim TV$



$Sales \sim Radio$

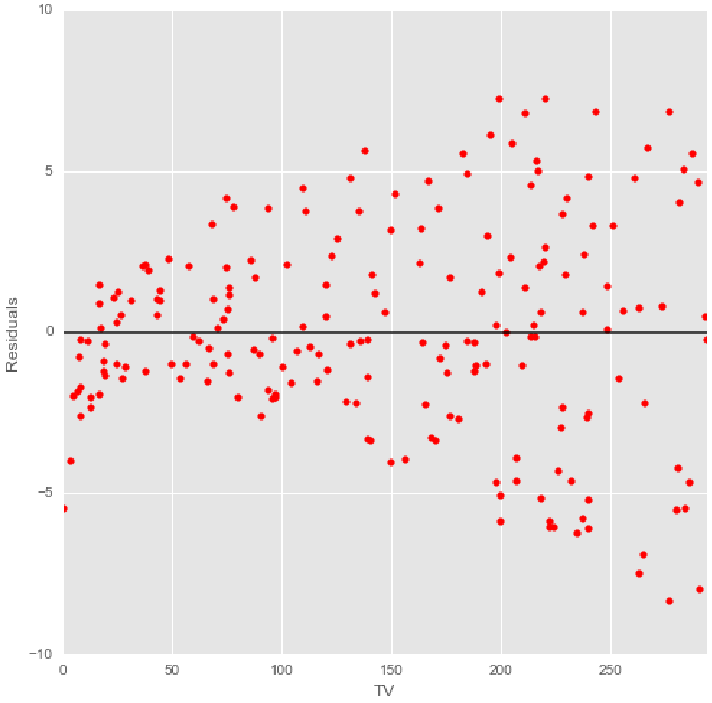


$Sales \sim Newspaper$

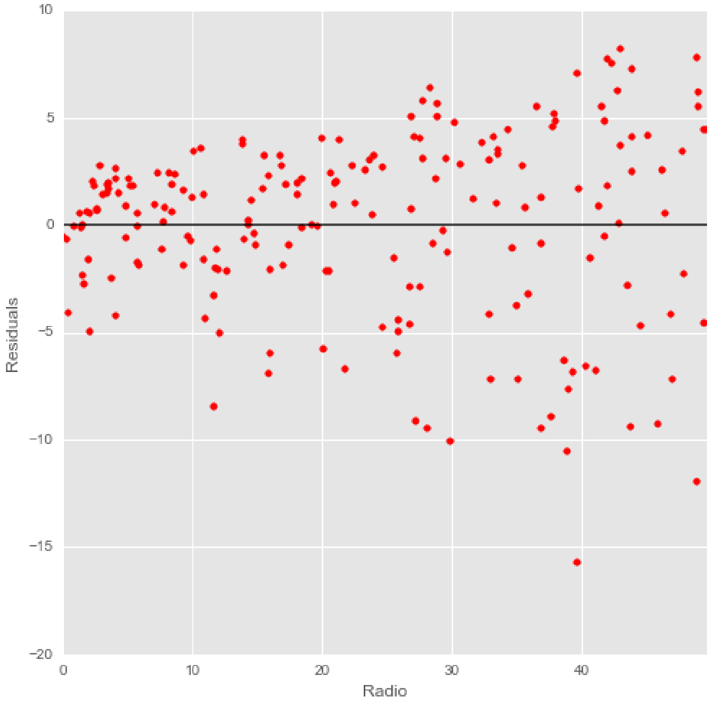


Scatterplots of residuals against advertising budget. Are they randomly distributed?

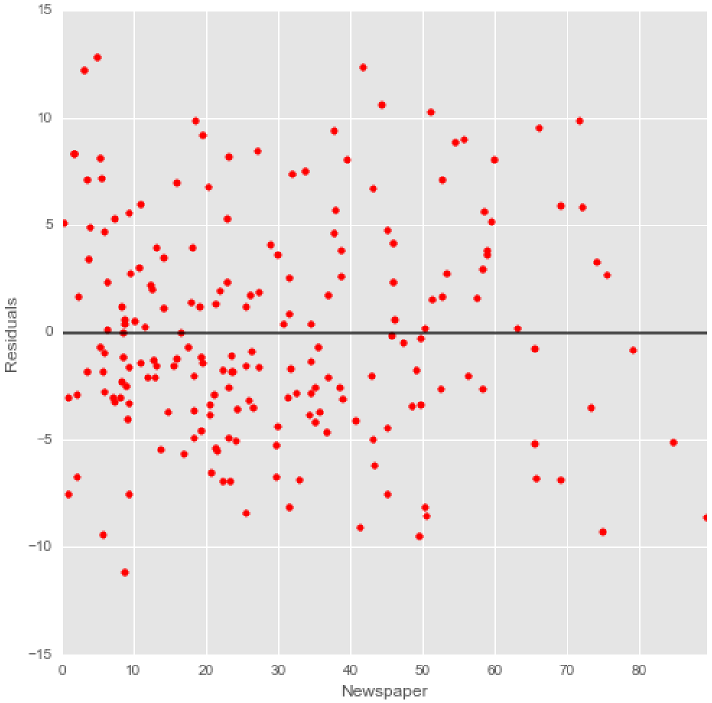
$Sales \sim TV$



$Sales \sim Radio$



$Sales \sim Newspaper$



$$Sales \sim TV + Radio + Newspaper$$

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?

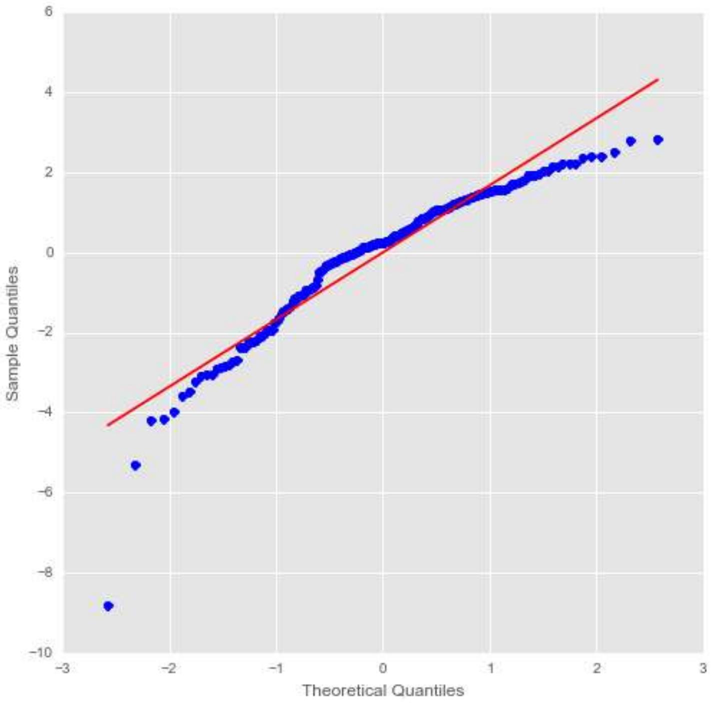
Dep. Variable:	Sales	**R-squared:**	0.895
Model:	OLS	**Adj. R-squared:**	0.894
Method:	Least Squares	**F-statistic:**	834.4
Date:		**Prob (F-statistic):**	2.60e-96
Time:		**Log-Likelihood:**	-383.26
No. Observations:	198	**AIC:**	772.5
Df Residuals:	195	**BIC:**	782.4
Df Model:	2		
Covariance Type:	nonrobust		

	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

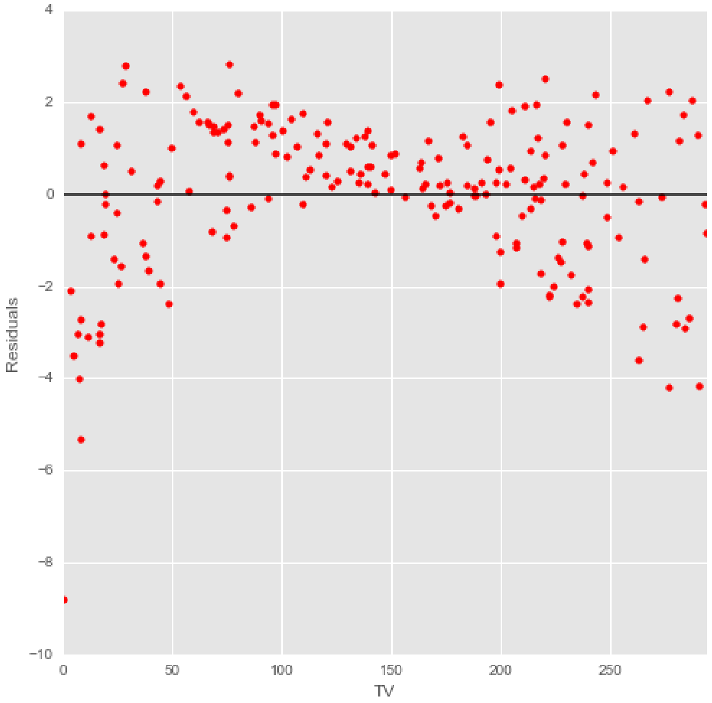
Omnibus:	59.228	**Durbin-Watson:**	2.038
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	145.127
Skew:	-1.321	**Prob(JB):**	3.06e-32
Kurtosis:	6.257	**Cond. No.**	423.

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

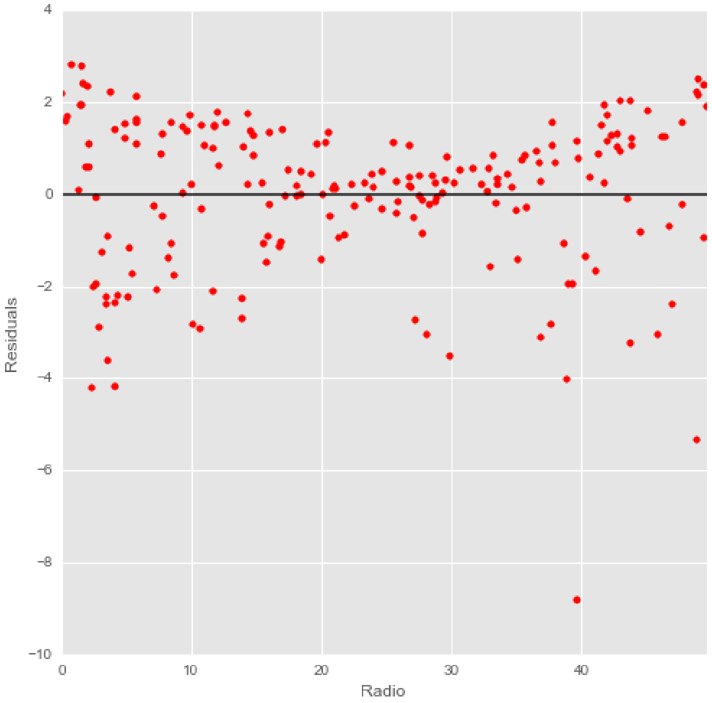
Residuals q-q plot



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 ($1,000) in TV is always ($\underbrace{.0457}_{\widehat{\beta}_1} \times \underbrace{.\$1{,}000}_{TV}$ = $45.7), 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 \$100,000, 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.)

- $Sales = \underbrace{6.76}_{\hat{\beta}'_0} + \underbrace{.0190}_{\hat{\beta}'_1} \times TV + \underbrace{.0276}_{\hat{\beta}'_2} \times Radio + \underbrace{.0011}_{\hat{\beta}'_3} \times TV \times Radio$
- 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

Activity | Interaction effects

EXERCISE

DIRECTIONS (10 minutes)

1. Our TV budget is $50,000 that we consider increasing it by $5,000. What would be the corresponding increase in sales based on different levels of radio budget?
   a. Consider the model without interactions first

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

   b. Then consider the model with interactions

$$Sales = \underbrace{6.76}_{\hat{\beta}'_0} + \underbrace{.0190}_{\hat{\beta}'_1} \times TV + \underbrace{.0276}_{\hat{\beta}'_2} \times Radio + \underbrace{.0011}_{\hat{\beta}'_3} \times TV \times Radio$$

2. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

Activity | Interaction effects (cont.)

EXERCISE

Radio budget	Model without interactions $Sales = \underbrace{2.93}_{\widehat{\beta}_0} + \underbrace{.0457}_{\widehat{\beta}_1} \times TV + \underbrace{.188}_{\widehat{\beta}_2} \times Radio$	Model with interactions $Sales = \underbrace{6.76}_{\widehat{\beta}'_0} + \underbrace{.0190}_{\widehat{\beta}'_1} \times TV + \underbrace{.0276}_{\widehat{\beta}'_2} \times Radio + \underbrace{.0011}_{\widehat{\beta}'_3} \times TV \times Radio$
Formula		
$15,000		
$10,000		
$5,000		

Activity | Interaction effects (cont.)

EXERCISE

Radio budget	Model without interactions	Model with interactions
	$Sales = \underbrace{2.93}_{\hat{\beta}_0} + \underbrace{.0457}_{\hat{\beta}_1} \times TV + \underbrace{.188}_{\hat{\beta}_2} \times Radio$	$Sales = \underbrace{6.76}_{\hat{\beta}'_0} + \underbrace{.0190}_{\hat{\beta}'_1} \times TV + \underbrace{.0276}_{\hat{\beta}'_2} \times Radio + \underbrace{.0011}_{\hat{\beta}'_3} \times TV \times Radio$
Formula	$\underbrace{.0457}_{\hat{\beta}_1} \times \Delta TV$	
\$15,000		
\$10,000		
\$5,000		

Activity | Interaction effects (cont.)

EXERCISE

Radio budget	Model without interactions $Sales = \underbrace{2.93}_{\hat{\beta}_0} + \underbrace{.0457}_{\hat{\beta}_1} \times TV + \underbrace{.188}_{\hat{\beta}_2} \times Radio$	Model with interactions $Sales = \underbrace{6.76}_{\hat{\beta}'_0} + \underbrace{.0190}_{\hat{\beta}'_1} \times TV +$ $\underbrace{.0276}_{\hat{\beta}'_2} \times Radio + \underbrace{.0011}_{\hat{\beta}'_3} \times TV \times Radio$
Formula	$\underbrace{.0457}_{\hat{\beta}_1} \times \Delta TV$	
$\$15{,}000$	$.0457 \times 5 = .228 = \$229$	
$\$10{,}000$	$\$229$	
$\$5{,}000$	$\$229$	

Activity | Interaction effects (cont.)

EXERCISE

Radio budget	Model without interactions $Sales = \underbrace{2.93}_{\hat{\beta}_0} + \underbrace{.0457}_{\hat{\beta}_1} \times TV + \underbrace{.188}_{\hat{\beta}_2} \times Radio$	Model with interactions $Sales = \underbrace{6.76}_{\hat{\beta}'_0} + \underbrace{.0190}_{\hat{\beta}'_1} \times TV +$ $\underbrace{.0276}_{\hat{\beta}'_2} \times Radio + \underbrace{.0011}_{\hat{\beta}'_3} \times TV \times Radio$
Formula	$\underbrace{.0457}_{\hat{\beta}_1} \times \Delta TV$	$\left(\underbrace{.0190}_{\hat{\beta}'_1} + \underbrace{.0011}_{\hat{\beta}'_3} \times Radio\right) \times \Delta TV$
$\$15{,}000$	$.0457 \times 5 = .228 = \$229$	$(.0190 + .0011 \times 15) \times 5$ $= .178 = \$178$
$\$10{,}000$	$\$229$	$(.0190 + .0011 \times 10) \times 5$ $= .150 = \$150$
$\$5{,}000$	$\$229$	$(.0190 + .0011 \times 5) \times 5$ $= .123 = \$123$

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