

Geometry of Multiple Linear Regression

Ex: Restaurants in NYC

ZAGAT

NEW YORK CITY ▾Ratings & ReviewsListsBuzz

Marea ➔

Italian | Midtown

ZAGAT RATED	FOOD 27	DECOR 26	SERVICE 26	COST \$117
------------------------	------------	-------------	---------------	---------------

Ratings are out of 30. Key to ratings ?

Our Summary Review


"Prepare to be thrilled" at this "all-star" Italian on Central Park South, which presents chef Michael White's "brilliantly executed" seafood and housemade pastas in "lovely", "refined" surroundings tended by a staff that's "gracious and on-cue"; just "dress pretty" and "bring your bank manager" since "they're not shy with the pricing" – though lunch is a "more affordable" option.

240 Central Park South | [Map](#)
New York, New York 10019
212-582-5100
marea-nyc.com

Today's hours:
12:00pm - 11:00pm

[VIEW MENU](#)

[✉ EMAIL](#) [📅 31 ADD](#)

[RESERVATIONS](#) 

Ex: Restaurants in NYC

```
nyc
```

```
## # A tibble: 168 x 7
##   Case Restaurant Price Food Decor Service East
##   <dbl> <chr>      <dbl> <dbl> <dbl>    <dbl> <dbl>
## 1     1 1 Daniella Ristorante    43    22    18     20     0
## 2     2 2 Tello's Ristorante    32    20    19     19     0
## 3     3 3 Biricchino           34    21    13     18     0
## 4     4 4 Bottino                41    20    20     17     0
## 5     5 5 Da Umberto             54    24    19     21     0
## 6     6 6 Le Madri               52    22    22     21     0
## 7     7 7 Le Zie                 34    22    16     21     0
## 8     8 8 Pasticcio              34    20    18     21     1
## 9     9 9 Belluno                39    22    19     22     1
## 10    10 10 Cinque Terre          44    21    17     19     1
## # ... with 158 more rows
```

What is the unit of observation?

A restaurant

What determines the price of a meal?

Let's look at the relationship between price, food rating, and decor rating.

$$Price \sim Food + Decor$$

```
m1 <- lm(Price ~ Food + Decor, data = nyc)
```

Model 1: Food + Decor

```
summary(m1)
```

```
##
## Call:
## lm(formula = Price ~ Food + Decor, data = nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.945  -3.766  -0.153   3.701  18.757
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -24.500      4.723   -5.19  6.2e-07 ***
## Food           1.646      0.262    6.29  2.7e-09 ***
## Decor          1.882      0.192    9.81  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.79 on 165 degrees of freedom
## Multiple R-squared:  0.617,    Adjusted R-squared:  0.612
## F-statistic: 133 on 2 and 165 DF,  p-value: <2e-16
```

The geometry of regression models

The function for \hat{y} is . . .

- A *line* when you have one continuous x .
- *Parallel lines* when you have one continuous x_1 and one categorical x_2 .
- *Unrelated lines* when you have one continuous x_1 , one categorical x_2 , and an interaction term $x_1 : x_2$.

When you have two continuous predictors x_1, x_2 , then your mean function is . . .

a plane

3d plot

interactive code

Location, location, location

Does the price depend on where the restaurant is located in Manhattan?

$$Price \sim Food + Decor + East$$

nyc

```
## # A tibble: 168 x 7
##   Case Restaurant Price Food Decor Service East
##   <dbl> <chr>      <dbl> <dbl> <dbl>    <dbl> <dbl>
## 1     1 1 Daniella Ristorante    43    22    18      20     0
## 2     2 2 Tello's Ristorante    32    20    19      19     0
## 3     3 3 Biricchino           34    21    13      18     0
## 4     4 4 Bottino              41    20    20      17     0
## 5     5 5 Da Umberto           54    24    19      21     0
## 6     6 6 Le Madri             52    22    22      21     0
## 7     7 7 Le Zie               34    22    16      21     0
## 8     8 8 Pasticcio            34    20    18      21     1
## 9     9 9 Belluno              39    22    19      22     1
## 10    10 10 Cinque Terre         44    21    17      19     1
## # ... with 158 more rows
```


Model 2: Food + Decor + East

```
m2 <- lm(Price ~ Food + Decor + East, data = nyc)
summary(m2)
```

```
##
## Call:
## lm(formula = Price ~ Food + Decor + East, data = nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.045   -3.881    0.039    3.392   17.756
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -24.027      4.673   -5.14  7.7e-07 ***
## Food           1.536      0.263    5.84  2.8e-08 ***
## Decor          1.909      0.190   10.05 < 2e-16 ***
## East           2.067      0.932    2.22  0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.72 on 164 degrees of freedom
## Multiple R-squared:  0.628,    Adjusted R-squared:  0.621
## F-statistic: 92.2 on 3 and 164 DF,  p-value: <2e-16
```

The geometry of regression models

- When you have two continuous predictors x_1, x_2 , then your mean function is *a plane*.
- When you have two continuous predictors x_1, x_2 , and a categorical predictor x_3 , then your mean function represents *parallel planes*.

3d Plot

interactive code

The geometry of regression models

- When you have two continuous predictors x_1 , x_2 , then your mean function is *a plane*.
- When you have two continuous predictors x_1 , x_2 , and a categorical predictor x_3 , then your mean function represents *parallel planes*.
- When you add in interaction effects, the planes become *tilted*.

Model 3: Food + Decor + East + Decor:East

```
m3 <- lm(Price ~ Food + Decor + East + Decor:East, data = nyc)
summary(m3)
```

```
##
## Call:
## lm(formula = Price ~ Food + Decor + East + Decor:East, data = nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.785   -3.665    0.378    3.729   17.636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -29.397     6.377   -4.61  8.1e-06 ***
## Food           1.663     0.282    5.90  2.1e-08 ***
## Decor          2.070     0.230    9.01  5.4e-16 ***
## East          9.662     6.218    1.55   0.12
## Decor:East    -0.435     0.352   -1.24   0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.71 on 163 degrees of freedom
## Multiple R-squared:  0.631,    Adjusted R-squared:  0.622
## F-statistic: 69.8 on 4 and 163 DF,  p-value: <2e-16
```

3d plot

interactive code

Comparing Models

- The **East** term was significant in model 2, suggesting that there is a significant relationship between location and price.
- That term became nonsignificant when we allowed the slope of **Decor** to vary with location, and that difference in slopes was also nonsignificant.
- Notice that slope estimate for a given variable will almost *always* change depending on the other variables that are in the model.