STP 530
Lecture 8: Interaction Regression Models

Yi Zheng, Ph.D.
yi.isabel.zheng@asu.edu

Arizona State University
School of Mathematical and Statistical Sciences

October 24, 2023

Outline

Additive models vs. Interaction models

Additive regression models:

$$E\{Y\} = \beta_0 + \underbrace{\beta_1 X_1}_{f(X_1)} + \underbrace{\beta_2 X_2}_{f(X_2)}$$

$$E\{Y\} = \beta_0 + \underbrace{\beta_1 X_1 + \beta_2 X_1^2}_{f(X_1)} + \underbrace{\beta_3 X_2}_{f(X_2)}$$

Note $f(X_1)$ is a function of solely X_1, and $f(X_2)$ is a function of solely X_2. The relation between $f(X_1)$ and $f(X_2)$ is additive. Such models are called **additive regression models**.

Additive models vs. Interaction models

Interaction regression models:

$$E\{Y\} = \beta_0 + \underbrace{\beta_1 X_1}_{\text{main effect of } X_1} + \underbrace{\beta_2 X_2}_{\text{main effect of } X_2} + \underbrace{\beta_3 X_1 X_2}_{\text{interaction effect}}$$

This model cannot be expressed in the form of additive functions of single predictors. So this model is not additive, and we call it an **interaction regression model**.

Interaction regression models

Examples of interaction regression models:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3$$

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3$$

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2$$

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$$

Outline

Example 1: Interaction effect of a numeric predictor and a categorical predictor

The NELS (National Education Longitudinal Study) data. A nationally representative sample of about 25,000 8th graders in 1988. Students were followed throughout secondary and postsecondary years.
`https://nces.ed.gov/surveys/nels88/`

In this example, we will use the wave 2 data (10th grade) to look at:

1. Does the measured self-esteem score differ between boys and girls?
2. Does the gender difference still exist after accounting for previous academic achievements?
3. Does the relationship between previous academic achievements and self-esteem differ for the two genders?

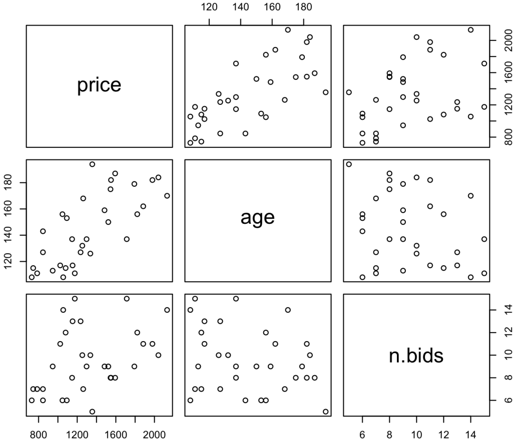
Example 2: Interaction effect of two numeric predictors

A collector of antique grandfather clocks sold at auction believes that the price received for the clocks depends on both the age of the clocks and the number of bidders at the auction.

$$y = \text{Auction price (dollars)}$$
$$x_1 = \text{Age of clock (years)}$$
$$x_2 = \text{Number of bidders}$$

Age, x_1	Number of Bidders, x_2	Auction Price, y	Age, x_1	Number of Bidders, x_2	Auction Price, y
127	13	$1,235	170	14	$2,131
115	12	1,080	182	8	1,550
127	7	845	162	11	1,884
150	9	1,522	184	10	2,041
156	6	1,047	143	6	845
182	11	1,979	159	9	1,483
156	12	1,822	108	14	1,055
132	10	1,253	175	8	1,545
137	9	1,297	108	6	729
113	9	946	179	9	1,792
137	15	1,713	111	15	1,175
117	11	1,024	187	8	1,593
137	8	1,147	111	7	785
153	6	1,092	115	7	744
117	13	1,152	194	5	1,356
126	10	1,336	168	7	1,262

Scatter plot matrix: `pairs(clock)`



- Moderate linear relationship between Y and each X.
- No associate between X_1 and X_2 — No multicollinearity seen here.

Fit the interaction model. Multicollinearity is found.

```
> m1 <- lm(price ~ age + n.bids + age:n.bids)
> summary(m1)

Call:
lm(formula = price ~ age + n.bids + age:n.bids)

Residuals:
     Min       1Q   Median       3Q      Max
-154.995  -70.431    2.069   47.880  202.259

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 320.4580   295.1413   1.086  0.28684
age           0.8781     2.0322   0.432  0.66896
n.bids      -93.2648    29.8916  -3.120  0.00416 **
age:n.bids    1.2978     0.2123   6.112 1.35e-06 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 88.91 on 28 degrees of freedom
Multiple R-squared:  0.9539,	Adjusted R-squared:  0.9489
F-statistic:   193 on 3 and 28 DF,  p-value: < 2.2e-16

> vif(m1) # Found very big vif values
       age     n.bids age:n.bids
  12.15313   28.25135   30.45770
```

Clock example continued

To reduce multicollinearity introduced by the interaction (product) term, we center the two predictors:

$$\text{age.c} = \text{age} - 145$$
$$\text{nbids.c} = \text{nbids} - 10$$

Note we are using the rounded integers instead of the precise means. The relief of multicollinearity is slightly lessened, but it helps with interpretability.

The fitted interaction model:

$$\widehat{\text{price}} = 1397 + 13.9(\text{age.c}) + 94.9(\text{nbids.c}) + 1.3(\text{age.c})(\text{nbids.c})$$

```
Call:
lm(formula = price ~ age.c + n.bids.c + age.c:n.bids.c)

Residuals:
     Min       1Q   Median       3Q      Max 
-154.995  -70.431    2.069   47.880  202.259 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1397.0169    16.6473  83.918  < 2e-16 ***
age.c            13.8566     0.6297  22.005  < 2e-16 ***
n.bids.c         94.9228     5.9964  15.830 1.69e-15 ***
age.c:n.bids.c    1.2978     0.2123   6.112 1.35e-06 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 88.91 on 28 degrees of freedom
Multiple R-squared:  0.9539,	Adjusted R-squared:  0.9489 
F-statistic:   193 on 3 and 28 DF,  p-value: < 2.2e-16

> vif(m2)
         age.c       n.bids.c age.c:n.bids.c 
      1.166929       1.136912       1.124721 
```

The standard errors of the first order terms have decreased noticeably, which indicates that centering did alleviate multicollinearity and recovered the previously inflated standard errors of the model coefficients.

The VIF values are now close to 1.

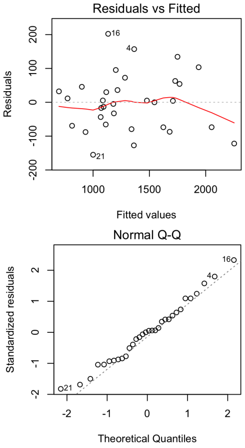
Compare and contrast: additive vs. interaction

```
> m3 <- lm(price ~ age.c + n.bids.c)
> summary(m3)

Call:
lm(formula = price ~ age.c + n.bids.c)

Residuals:
    Min      1Q  Median      3Q     Max
-206.49 -117.34   16.66  102.55  213.50

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1367.9617    23.9516  57.114  < 2e-16 ***
age.c         12.7406     0.9047  14.082 1.69e-14 ***
n.bids.c      85.9530     8.7285   9.847 9.34e-11 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 133.5 on 29 degrees of freedom
Multiple R-squared:  0.8923,	Adjusted R-squared:  0.8849
F-statistic: 120.2 on 2 and 29 DF,  p-value: 9.216e-15
```

```
> m2 <- lm(price ~ age.c + n.bids.c + age.c:n.bids.c)
> summary(m2)

Call:
lm(formula = price ~ age.c + n.bids.c + age.c:n.bids.c)

Residuals:
     Min       1Q   Median       3Q      Max
-154.995  -70.431    2.069   47.880  202.259

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1397.0169    16.6473  83.918  < 2e-16 ***
age.c            13.8566     0.6297  22.005  < 2e-16 ***
n.bids.c         94.9228     5.9964  15.830 1.69e-15 ***
age.c:n.bids.c    1.2978     0.2123   6.112 1.35e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 88.91 on 28 degrees of freedom
Multiple R-squared:  0.9539,	Adjusted R-squared:  0.9489
F-statistic:   193 on 3 and 28 DF,  p-value: < 2.2e-16
```

Test whether interaction should be retained (Run diagnostics first)

```
> m2 <- lm(price ~ age.c + n.bids.c + age.c:n.bids.c)
> m3 <- lm(price ~ age.c + n.bids.c)
>
> par(mfrow=c(2, 2))
> plot(m2)
```




```
> # Statistical significance: the general linear test approach
>
> anova(m2, m3)
Analysis of Variance Table

Model 1: price ~ age.c + n.bids.c + age.c:n.bids.c
Model 2: price ~ age.c + n.bids.c
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     28 221362
2     29 516727 -1   -295364 37.361 1.353e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Practical significance:
>
> # Compare the MS of the interaction against the main effect MSs
> anova(m2)
Analysis of Variance Table

Response: price
               Df  Sum Sq Mean Sq F value    Pr(>F)
age.c           1 2555224 2555224 323.209 < 2.2e-16 ***
n.bids.c        1 1727838 1727838 218.554 9.382e-15 ***
age.c:n.bids.c  1  295364  295364  37.361 1.353e-06 ***
Residuals      28  221362    7906
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # The additive model R2
> summary(m3)$r.squared
[1] 0.8923439
>
> # The interaction model R2
> summary(m2)$r.squared
[1] 0.9538809
>
> # The partial R2 for the interaction term: R2_{Y,int|age,nbids}
> rsq.partial(objF=m2, objR=m3)$partial.rsq
[1] 0.5716068
```

Interpretation of model coefficients

The fitted interaction model:

$$\widehat{\text{price}} = 1397 + 13.9(\text{age.c}) + 94.9(\text{nbids.c}) + 1.3(\text{age.c})(\text{nbids.c})$$

Recall in additive models:

Slope b_1: Holding all other predictors constant, for every 1-unit increase of X_1, the expected Y value changes by b_1.

Does this interpretation still work for interaction models?

In the **additive model** $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

The change in $E\{Y\}$ with one-unit increase in X_1, holding X_2 constant is:

$$\frac{\partial E\{Y\}}{\partial X_1} = \beta_1$$

In the **interaction model** $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

The change in $E\{Y\}$ with one-unit increase in X_1, holding X_2 constant is:

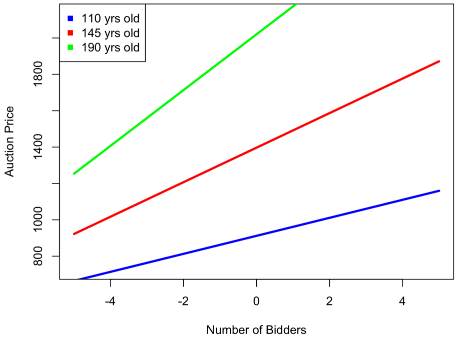
$$\frac{\partial E\{Y\}}{\partial X_1} = \beta_1 + \beta_3 X_2$$

The change in $E\{Y\}$ with one-unit increase in X_2, holding X_1 constant is:

$$\frac{\partial E\{Y\}}{\partial X_2} = \beta_2 + \beta_3 X_1$$

The fitted interaction model:

$$\widehat{\text{price}} = 1397 + 13.9(\text{age.c}) + 94.9(\text{nbids.c}) + 1.3(\text{age.c})(\text{nbids.c})$$



- The effect of number of bidders on auction price depends on the age of the clock being auctioned (vice versa).
- The regression line of auction price on number of bidders has **different slopes** given different ages of the clock (**non-parallel**).

Outline

Additive vs. interaction regression

1. When the relationship between the response variable and a predictor **DOES NOT depend on** the values of another predictor in the model:
    - There is **no interaction** between the effects of the two predictors on the response variable (a.k.a., **additive regression models**).
    - The regression lines of Y on one predictor given different values of the other predictor are a set of **parallel** straight lines.
2. When the relationship between the response variable and a predictor **DOES depend on** the values of another predictor in the model
    - There is an **interaction** between the effects of the two predictors on the response variable.
    - The regression lines of Y on one predictor given different values of the other predictor are **not parallel**.

Outline

Reinforcement interaction vs. interference interaction

Assume both β_1 and β_2 are **positive** in the interaction regression model below

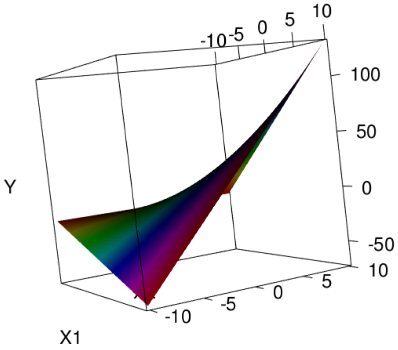
$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2,$$

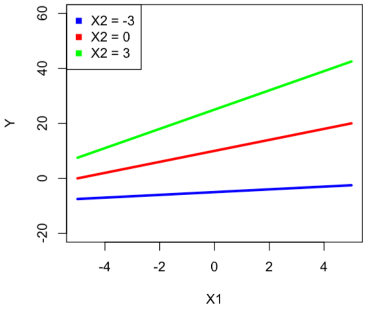
- A **positive** coefficient for the interaction term ($\beta_3 > 0$): **Reinforcement** interaction effect
- A **negative** coefficient for the interaction term ($\beta_3 < 0$): **Interference** interaction effect

Note: When β_1 and β_2 are **negative**, a negative β_3 is usually viewed as a reinforcement type and a positive β_3 is viewed as an interference type.

Reinforcement interaction effect

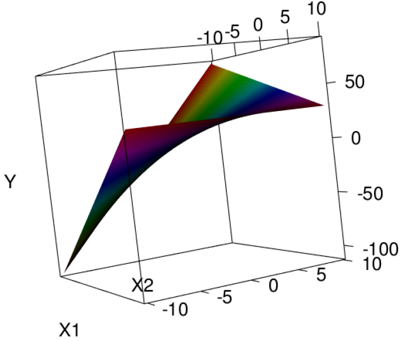
$$E\{Y\} = 1 + 2X_1 + 5X_2 + .5X_1X_2 \qquad \text{(Textbook Equation 8.26, p.307)}$$

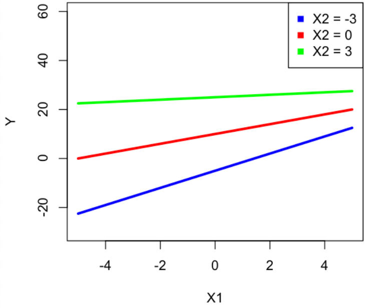




Interference interaction effect

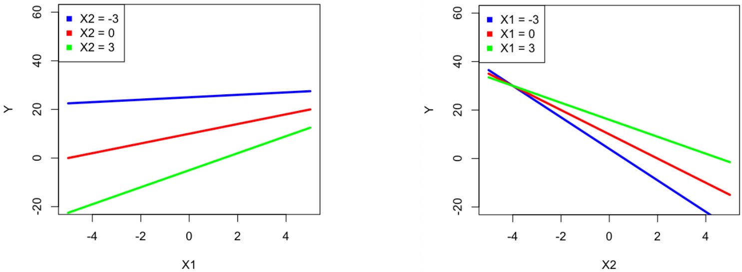
$$E\{Y\} = 1 + 2X_1 + 5X_2 - .5X_1X_2 \qquad \text{(Textbook Equation 8.27, p.308)}$$





What about other scenarios of the signs for β_1 and β_2??

$$E\{Y\} = 1 + 2X_1 - 5X_2 + .5X_1X_2$$

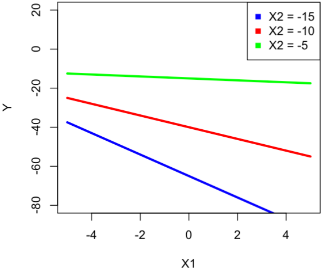


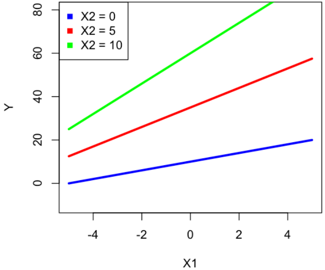
In fact, whether "reinforcing" or "interfering" is specific to the individual predictor:

- Interaction coefficient has the **same sign** as the main effect coefficient: **Reinforcement** interaction effect
- Interaction coefficient has the **opposite sign** as the main effect coefficient: **Interference** interaction effect

This also depends on the specific values of **the other variable**. An interference interaction effect can turn into a reinforcement interaction. Vice versa.

$$E\{Y\} = 1 + 2X_1 + 5X_2 + .5X_1X_2$$





In summary, the nature of the interaction depends on:

1. The specific effect of focus (the focus predictor); and
2. The range of the other variable being considered for the research.

Outline

Express the fitted model with the original (uncentered) predictors

$$
\begin{aligned}
\widehat{\text{price}} =& 1397 + 13.9(\text{age.c}) + 94.9(\text{nbids.c}) + 1.3(\text{age.c})(\text{nbids.c}) \\
=& 1397 + 13.9(\text{age} - 145) + 94.9(\text{nbids} - 10) \\
& + 1.3(\text{age} - 145)(\text{nbids} - 10) \\
=& 1397 + 13.9(\text{age}) - 13.9 * 145 + 94.9(\text{nbids}) - 94.9 * 10 \\
& + 1.3(\text{age})(\text{nbids}) - 1.3 * 10 * (\text{age}) - 1.3 * 145 * (\text{nbids}) \\
& + 1.3 * 10 * 145 \\
=& 317.5 + 0.9(\text{age}) - 93.6(\text{nbids}) + 1.3(\text{age})(\text{nbids})
\end{aligned}
$$

Compare the recovered coefficients with those obtained by directly fitting the uncentered data

```
> b0.c <- coef(m2)[1]
> b1.c <- coef(m2)[2]
> b2.c <- coef(m2)[3]
> b3.c <- coef(m2)[4]
>
> c(b0.c, b1.c, b2.c, b3.c)
   (Intercept)         age.c      n.bids.c age.c:n.bids.c
   1397.016853     13.856601     94.922820       1.297846
>
> b0 <- b0.c - b1.c * 145 - b2.c * 10 + b3.c * 145 * 10
> b1 <- b1.c - b3.c * 10
> b2 <- b2.c - b3.c * 145
> b3 <- b3.c
>
> # Coefficients recovered from the centered model
> as.numeric(c(b0, b1, b2, b3))
[1] 320.4579934   0.8781425 -93.2648244   1.2978458
>
> # Coefficients obtained from fitting the uncentered predictors
> coef(m1)
(Intercept)         age      n.bids   age:n.bids
320.4579934   0.8781425 -93.2648244   1.2978458
```

Compare and contrast: uncentered vs. centered

```
> m1 <- lm(price ~ age + n.bids + age:n.bids)
> summary(m1)

Call:
lm(formula = price ~ age + n.bids + age:n.bids)

Residuals:
     Min       1Q   Median       3Q      Max 
-154.995  -70.431    2.069   47.880  202.259 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 320.4580   295.1413   1.086  0.28684    
age           0.8781     2.0322   0.432  0.66896    
n.bids      -93.2648    29.8916  -3.120  0.00416 ** 
age:n.bids    1.2978     0.2123   6.112 1.35e-06 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 88.91 on 28 degrees of freedom
Multiple R-squared:  0.9539,	Adjusted R-squared:  0.9489 
F-statistic:   193 on 3 and 28 DF,  p-value: < 2.2e-16

> vif(m1) # Found very big vif values
     age   n.bids age:n.bids 
12.15313 28.25135 30.45770 
```

```
> m2 <- lm(price ~ age.c + n.bids.c + age.c:n.bids.c)
> summary(m2)

Call:
lm(formula = price ~ age.c + n.bids.c + age.c:n.bids.c)

Residuals:
     Min       1Q   Median       3Q      Max 
-154.995  -70.431    2.069   47.880  202.259 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1397.0169    16.6473  83.918  < 2e-16 ***
age.c            13.8566     0.6297  22.005  < 2e-16 ***
n.bids.c         94.9228     5.9964  15.830 1.69e-15 ***
age.c:n.bids.c    1.2978     0.2123   6.112 1.35e-06 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 88.91 on 28 degrees of freedom
Multiple R-squared:  0.9539,	Adjusted R-squared:  0.9489 
F-statistic:   193 on 3 and 28 DF,  p-value: < 2.2e-16

> vif(m2)
         age.c       n.bids.c age.c:n.bids.c 
      1.166929       1.136912       1.124721 
```

Outline

Tips of using interaction regression

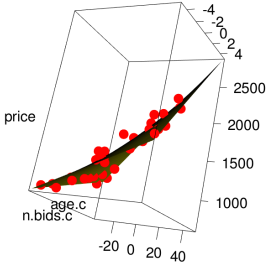
- Even if the predictors themselves are uncorrelated, the interaction term (the product term) will likely introduce multicollinearity to the model. You can alleviate the situation by centering the predictors.
- When there are k possible predictors, the full first-order interaction model includes 2^k terms, which makes the model too large to interpret. It is therefore desirable to identify in advance those interactions that are most likely to influence the response variable in important ways.
- One can also plot the residuals of an additive model against different product terms to determine which ones appear to be significant.
- If you decide to retain one interaction term, all main effect terms of the predictors involved in this interaction term should also be retained in the model.

Interaction models without main effect terms

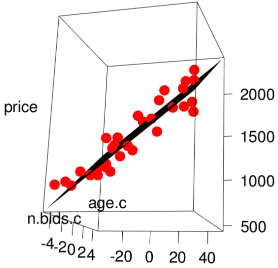
```
> m4 <- lm(price ~ age.c + age.c:n.bids.c)
> summary(m4)

Call:
lm(formula = price ~ age.c + age.c:n.bids.c)

Residuals:
    Min      1Q  Median      3Q     Max 
-473.82 -244.19   23.07  192.45  481.80 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1336.6247    50.2238  26.613  < 2e-16 ***
age.c            10.8021     1.8578   5.814 2.65e-06 ***
age.c:n.bids.c    0.4753     0.6381   0.745    0.462    
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 275.6 on 29 degrees of freedom
Multiple R-squared:  0.5411, Adjusted R-squared:  0.5095 
F-statistic:  17.1 on 2 and 29 DF,  p-value: 1.243e-05
```

```
> m5 <- lm(price ~ age.c:n.bids.c)
> summary(m5)

Call:
lm(formula = price ~ age.c:n.bids.c)

Residuals:
    Min      1Q  Median      3Q     Max 
-540.45 -293.75  -75.37  211.88  850.44 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1319.4500    72.5440  18.188   <2e-16 ***
age.c:n.bids.c   -0.3889     0.8979  -0.433    0.668    
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 398.7 on 30 degrees of freedom
Multiple R-squared:  0.006214,   Adjusted R-squared:  
-0.02691 
F-statistic: 0.1876 on 1 and 30 DF,  p-value: 0.668
```

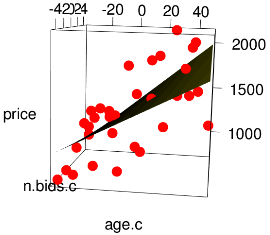
Model 2 (with interaction)



Model 3 (no interaction)



Model 4 (no nbids main effect)



Model 5 (none main effects)