STAT 420

A collector of antique grandfather clocks knows that the price received for the clocks increases linearly with the age of the clocks. Moreover, the collector hypothesizes that the auction price of the clocks will increase linearly as the number of bidders increases. Thus, the following first-order model is hypothesized:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where

Y = Auction price (dollars)

x_1 = Age of clock (years)

x_2 = Number of bidders

A sample of 32 auction prices of grandfather clocks, along with their age and the number of bidders, is given in the table below.

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

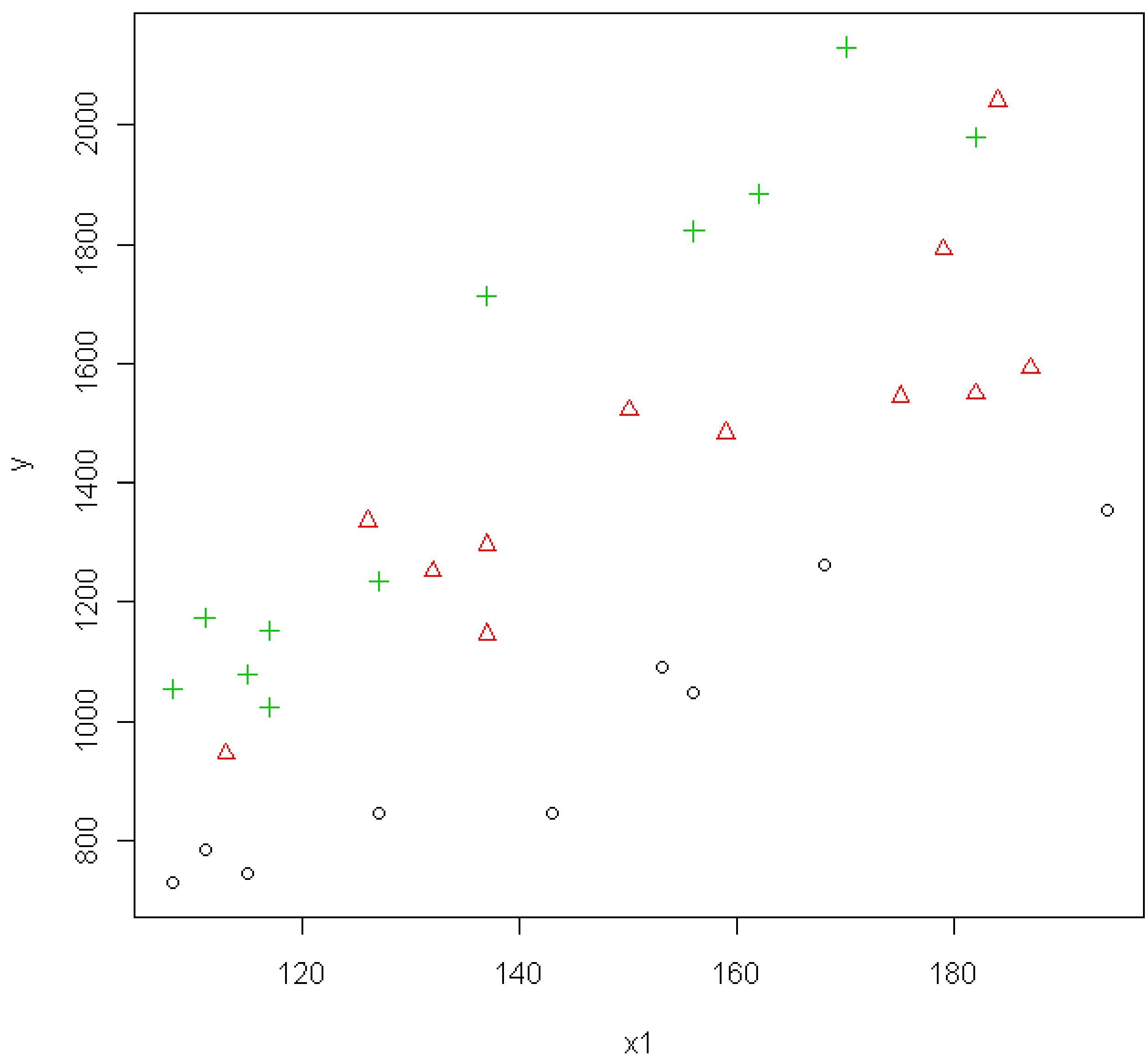
```
> clocks.dat = read.table(" ... /clocks.csv",sep=",",header=T)
>
> clocks.fit1 = lm(y ~ x1 + x2, data=clocks.dat)
> summary(clocks.fit1)

Call:
lm(formula = y ~ x1 + x2, data = clocks.dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-206.48 -117.34   16.66  102.55  213.50 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1338.9513   173.8095  -7.704 1.71e-08 ***
x1             12.7406     0.9047  14.082 1.69e-14 ***
x2             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
```

Suppose the collector of grandfather clocks, having observed many auctions, believes that the *rate of increase* of the auction price with age will be driven upward by a large number of bidders. Consequently, the interaction model is proposed:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

```
> clocks.dat$x1x2 = clocks.dat$x1 * clocks.dat$x2
> clocks.fit2 = lm(y ~ x1 + x2 + x1x2, data=clocks.dat)
> summary(clocks.fit2)

Call:
lm(formula = y ~ x1 + x2 + x1x2, data = clocks.dat)

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    
x1            0.8781     2.0322   0.432  0.66896    
x2          -93.2648    29.8916  -3.120  0.00416 ** 
x1x2          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
```

```
> gr = rep(1,32)
>
> for (i in 1:32) {
+ if (x2[i]>7) {gr[i]=gr[i]+1}
+ if (x2[i]>10) {gr[i]=gr[i]+1}
+ }
>
> plot(x1,y,pch=gr,col=gr)
```



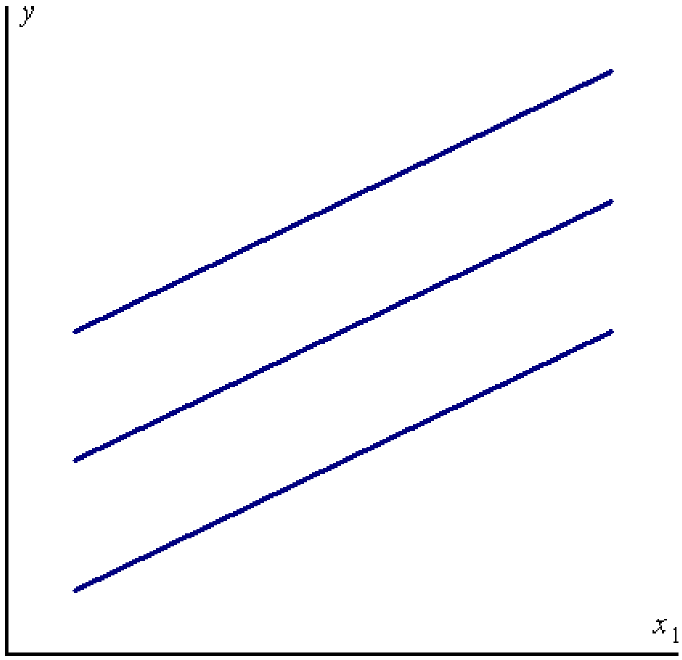
An Interaction Model Relating Y to Two Quantitative Independent Variables

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

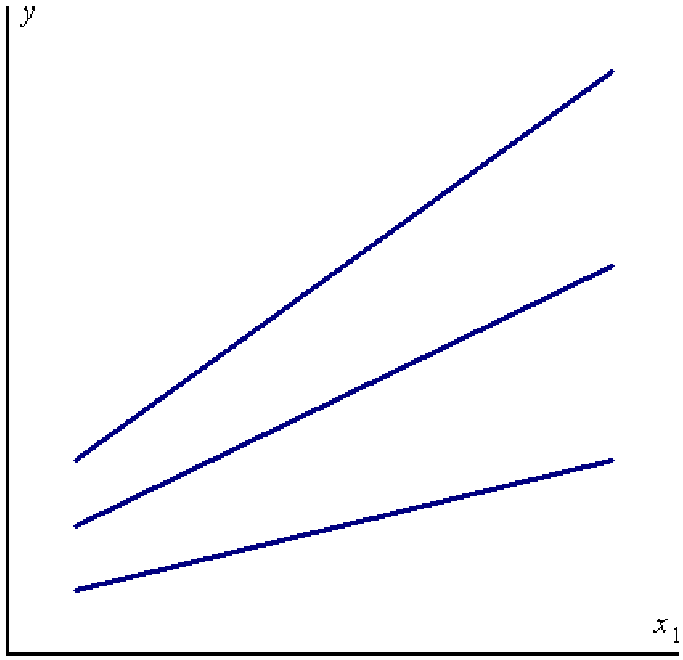
where

$(\beta_1 + \beta_3 x_2)$ represents the change in $E(Y)$ for every 1-unit increase in x_1, holding x_2 fixed.

$(\beta_2 + \beta_3 x_1)$ represents the change in $E(Y)$ for every 1-unit increase in x_2, holding x_1 fixed.



No interaction between x_1 and x_2



Interaction between x_1 and x_2

Estimate the change in auction price of a 170-year-old grandfather clock, y, for each additional bidder.

$$\hat{\beta}_2 \approx 85.95$$

We estimate that the auction price of a grandfather clock will increase by about \$85.95 for every additional bidder.

$$\hat{\beta}_2 + \hat{\beta}_3 x_1 \approx -93.265 + 1.2978 \cdot 170 \approx 127.36$$

We estimate that the auction price of a 170-year-old clock will increase by about \$127.36 for every additional bidder.

```
> clocks.dat$y[17] = 1131
```
– **creating an outlier**

```
> clocks.fit3 = lm(y ~ x1 + x2 + x1x2, data=clocks.dat)
> summary(clocks.fit3)

Call:
lm(formula = y ~ x1 + x2 + x1x2, data = clocks.dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-783.43  -75.51  -11.01  125.33  285.70 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) -512.8102   665.8601  -0.770   0.4477  
x1             8.1651     4.5847   1.781   0.0858 .
x2            19.8877    67.4376   0.295   0.7702  
x1x2           0.3196     0.4790   0.667   0.5101  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 200.6 on 28 degrees of freedom
Multiple R-Squared: 0.7292,     Adjusted R-squared: 0.7002 
F-statistic: 25.13 on 3 and 28 DF,  p-value: 4.277e-08
```

```
> X = cbind(rep(1,32),clocks.dat$x1,clocks.dat$x2,clocks.dat$x1*clocks.dat$x2)
>
> ## the hat-matrix
> H = X %*% solve(t(X)%*%X) %*% t(X)
>
> ##  leverages
> lev = rep(0,32)
> for (i in 1:32) {lev[i] = H[i,i]}
> lev
 [1] 0.08744552 0.09588217 0.09284425 0.03427419 0.09043677 0.17071980
 [7] 0.09518823 0.03923118 0.03787039 0.08540308 0.15283620 0.07005945
[13] 0.04868743 0.08641369 0.12081830 0.04687672 0.33847490 0.09805986
[19] 0.08064726 0.12424110 0.08628552 0.04090267 0.23824594 0.07618809
[25] 0.24675474 0.08234239 0.28575991 0.11688367 0.16245986 0.14159717
[31] 0.44331534 0.08285421
```

OR

```
> influence(clocks.fit2)$hat
         1          2          3          4          5          6          7 
0.08744552 0.09588217 0.09284425 0.03427419 0.09043677 0.17071980 0.09518823 
         8          9         10         11         12         13         14 
0.03923118 0.03787039 0.08540308 0.15283620 0.07005945 0.04868743 0.08641369 
        15         16         17         18         19         20         21 
0.12081830 0.04687672 0.33847490 0.09805986 0.08064726 0.12424110 0.08628552 
        22         23         24         25         26         27         28 
0.04090267 0.23824594 0.07618809 0.24675474 0.08234239 0.28575991 0.11688367 
        29         30         31         32 
0.16245986 0.14159717 0.44331534 0.08285421 
```

```
> sum(lev)
[1] 4
```