

```
#####Insurance innovation example#####
```

```
> ex.data
```

```
  y size  type
1 17 151 mutual
2 26  92 mutual
3 21 175 mutual
4 30  31 mutual
5 22 104 mutual
6  0 277 mutual
7 12 210 mutual
8 19 120 mutual
9  4 290 mutual
10 16 238 mutual
11 28 164 stock
12 15 272 stock
13 11 295 stock
14 38  68 stock
15 31  85 stock
16 21 224 stock
17 20 166 stock
18 13 305 stock
19 30 124 stock
20 14 246 stock
```

```
myfit1<-lm(y~size+type)
```

```
#as a default, mutual is treated as 0 and stock is treated as 1
```

```
myfit1
```

```
summary(myfit1)
```

```
> summary(myfit1)
```

```
Call:
```

```
lm(formula = y ~ size + type)
```

```
Residuals:
```

```
  Min   1Q  Median   3Q   Max
-5.6915 -1.7036 -0.4385  1.9210  6.3406
```

```
Coefficients:
```

```
            Estimate    Std. Error  t value    Pr(>|t|)
(Intercept) 33.874069   1.813858  18.675   9.15e-13 ***
size        -0.101742   0.008891  -11.443  2.07e-09 ***
typestock    8.055469   1.459106   5.521   3.74e-05 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.221 on 17 degrees of freedom
```

```
Multiple R-squared:  0.8951,    Adjusted R-squared:  0.8827
```

```
F-statistic: 72.5 on 2 and 17 DF, p-value: 4.765e-09
```

```

> #code "type" using indicator variable, x2=0, mutual, 1 for stock
> for(i in 1:n){
+ if (ex.data$type[i]=='mutual') {ex.data$x2[i]<-0}
+ if (ex.data$type[i]=='stock') {ex.data$x2[i]<-1}
+ }
> ex.data
  y size type x2
1 17 151 mutual 0
2 26  92 mutual 0
3 21 175 mutual 0
4 30  31 mutual 0
5 22 104 mutual 0
6  0 277 mutual 0
7 12 210 mutual 0
8 19 120 mutual 0
9  4 290 mutual 0
10 16 238 mutual 0
11 28 164 stock 1
12 15 272 stock 1
13 11 295 stock 1
14 38  68 stock 1
15 31  85 stock 1
16 21 224 stock 1
17 20 166 stock 1
18 13 305 stock 1
19 30 124 stock 1
20 14 246 stock 1

```

```
> summary(myfit2)
```

Call:

```
lm(formula = y ~ size + x2, data = ex.data)
```

Residuals:

```

  Min    1Q  Median    3Q   Max
-5.6915 -1.7036 -0.4385  1.9210  6.3406

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.874069   1.813858  18.675 9.15e-13 ***
size        -0.101742   0.008891  -11.443 2.07e-09 ***
x2           8.055469   1.459106   5.521 3.74e-05 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.221 on 17 degrees of freedom

Multiple R-squared: 0.8951, Adjusted R-squared: 0.8827

F-statistic: 72.5 on 2 and 17 DF, p-value: 4.765e-09

##fit with interaction term

```
> myfit3<-lm(y~size+type+size*type)
> summary(myfit3)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	33.8383695	2.4406498	13.864	2.47e-10 ***
size	-0.1015306	0.0130525	-7.779	7.97e-07 ***
typestock	8.1312501	3.6540517	2.225	0.0408 *
size:typestock	-0.0004171	0.0183312	-0.023	0.9821

---

Residual standard error: 3.32 on 16 degrees of freedom  
 Multiple R-squared: 0.8951, Adjusted R-squared: 0.8754  
 F-statistic: 45.49 on 3 and 16 DF, p-value: 4.675e-08

```
> myfit4<-lm(y~size+x2+size*x2,data=ex.data)
> summary(myfit4)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	33.8383695	2.4406498	13.864	2.47e-10 ***
size	-0.1015306	0.0130525	-7.779	7.97e-07 ***
x2	8.1312501	3.6540517	2.225	0.0408 *
size:x2	-0.0004171	0.0183312	-0.023	0.9821

---

Residual standard error: 3.32 on 16 degrees of freedom  
 Multiple R-squared: 0.8951, Adjusted R-squared: 0.8754  
 F-statistic: 45.49 on 3 and 16 DF, p-value: 4.675e-08

```
> anova(myfit4)
Analysis of Variance Table
```

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
size	1	1188.17	1188.17	107.7819	1.627e-08 ***
x2	1	316.25	316.25	28.6875	6.430e-05 ***
size:x2	1	0.01	0.01	0.0005	0.9821
Residuals	16	176.38	11.02		

```
> myfit0<-lm(y~size)
> myfit0
```

###test for  $\beta_2 = \beta_3 = 0$ , or compare model with size and model with  
 ##interaction

```
> anova(myfit0,myfit4)
```

Model 1: y ~ size

Model 2: y ~ size + x2 + size \* x2

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	18	492.63				
2	16	176.38	2	316.25	14.344	0.00027 ***

```
##reject the reduced model, conclude that the two regression functions are not
##identical
```

```
##three categories#####
```

```
>> ex.data2
  V1 V2 V3
1 17 151 0
2 26 92 0
3 21 175 0
4 30 31 0
5 22 104 0
6 0 277 0
7 12 210 1
8 19 120 1
9 4 290 1
10 16 238 1
11 28 164 1
12 15 272 1
13 11 295 3
14 38 68 3
15 31 85 3
16 21 224 3
17 20 166 3
18 13 305 3
19 30 124 3
20 14 246 3
```

```
> y<-ex.data2$V1
> x1<-ex.data2$V2
> x2<-ex.data2$V3
> f.x2<-factor(x2) #treat x2 as a categorical variable
> f.x2
[1] 0 0 0 0 0 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3
Levels: 0 1 3
> x2
[1] 0 0 0 0 0 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3
>
> myfit5<-lm(y~x1+x2) #x2 is treated as continuous variable
> myfit5
```

```
Call:
lm(formula = y ~ x1 + x2)
```

```
Coefficients:
(Intercept)      x1      x2
  33.9568   -0.1011   2.5503
```

```
> summary(myfit5)
```

```
Call:
```

```
lm(formula = y ~ x1 + x2)
```

```
Residuals:
```

```
  Min      1Q  Median      3Q      Max
-5.9563 -2.8538 -0.8054  2.4839  8.0708
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.95685   2.33834   14.522 5.17e-11 ***
x1           -0.10108   0.01134   -8.915 8.11e-08 ***
x2            2.55029   0.72432    3.521 0.00262 **
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.094 on 17 degrees of freedom
```

```
Multiple R-squared:  0.8305, Adjusted R-squared:  0.8106
```

```
F-statistic: 41.65 on 2 and 17 DF, p-value: 2.804e-07
```

```
>
```

```
> myfit6<-lm(y~x1+f.x2)
```

```
> myfit6
```

```
Call:
```

```
lm(formula = y ~ x1 + f.x2)
```

```
Coefficients:
```

```
(Intercept)      x1      f.x21      f.x23
  33.7317   -0.1041   4.3825   8.2033
```

```
#note that there are two coefficients associated with x2, since  
x2 has three categories. f.x21 is the coefficient associated with category 1, and  
f.x23 is the coefficient associated with category 3.
```

```
> summary(myfit6)
```

```
Call:
```

```
lm(formula = y ~ x1 + f.x2)
```

```
Residuals:
```

```
  Min      1Q  Median      3Q      Max
-6.6241 -2.7301 -0.3676  2.6961  6.9556
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.73168   2.38911   14.119 1.89e-10 ***
x1           -0.10408   0.01218   -8.545 2.33e-07 ***
f.x21         4.38253   2.57383    1.703 0.10796
f.x23         8.20329   2.32441    3.529 0.00279 **
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4.149 on 16 degrees of freedom  
Multiple R-squared: 0.8362, Adjusted R-squared: 0.8054  
F-statistic: 27.22 on 3 and 16 DF, p-value: 1.603e-06

Response function for  $x_2=0$ ,  $E(y) = 33.73 - 0.10408 x_1$

Response function for  $x_2=1$ ,  $E(y) = (33.73+4.38253) - 0.10408 x_1$

Response function for  $x_2=3$ ,  $E(y) = (33.73+8.20329) - 0.10408 x_1$