Stat 206: Linear Models

Lecture 13

Nov. 13, 2019

Polynomial Regression

Polynomial regression models are among the most commonly used models to describe a regression relation.

- Polynomial regression models are very flexible and are easy to fit.
- Polynomial models with higher than third-order terms are rarely employed in practice.
  - They often lead to estimators.
  - They might fit the observed data , but generalize well to new observations, a phenomena called .

Polynomial Regression

Polynomial regression models are among the most commonly used models to describe a curvilinear regression relation.

- Polynomial regression models are very flexible and are easy to fit.
- Polynomial models with higher than third-order terms are rarely employed in practice.
    - They often lead to highly variable estimators.
    - They might fit the observed data well, but could fail to generalize well to new observations, a phenomena called *overfitting*.

Second-Order Model with One Predictor

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1(X_i - \overline{X}) + \beta_2(X_i - \overline{X})^2 + \epsilon_i \\
&= \beta_0 + \beta_1\tilde{X}_i + \beta_2\tilde{X}_i^2 + \epsilon_i, \quad i = 1, \cdots, n,
\end{aligned}
$$

where $\tilde{X}_i = X_i - \overline{X}$ is the centered value of the predictor variable in the ith case.

- Centering often between the linear term X and the quadratic term X^2 substantially *(Why?)* and thus improves numerical accuracy. *Will centering change the fitted regression function?*
- The response function is a parabola:
  - β_0 is the mean response when
  .
  - β_1 is called the and β_2 is called the
  .

Second-Order Model with One Predictor

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1(X_i - \overline{X}) + \beta_2(X_i - \overline{X})^2 + \epsilon_i \\
&= \beta_0 + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \epsilon_i, \quad i = 1, \cdots, n,
\end{aligned}
$$

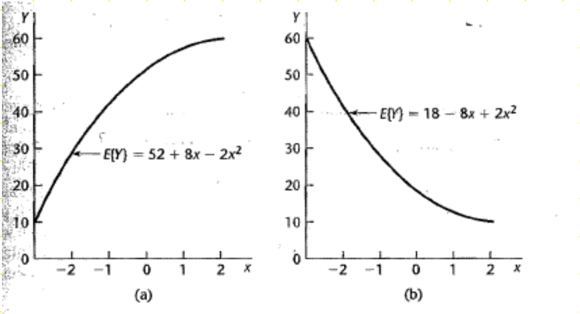
where $\tilde{X}_i = X_i - \overline{X}$ is the centered value of the predictor variable in the ith case.

- Centering often reduces the correlation between the linear term X and the quadratic term X^2 substantially *(Why?)* and thus improves numerical accuracy. *Will centering change the fitted regression function?*
- The response function is a parabola:

$$
\begin{aligned}
E(Y) &= \beta_0 + \beta_1(X - \overline{X}) + \beta_2(X - \overline{X})^2 \\
&= \beta_0 + \beta_1 \tilde{X} + \beta_2 \tilde{X}^2.
\end{aligned}
$$

  - β_0 is the mean response when $\tilde{X} = 0$, i.e. when $X = \overline{X}$.
  - β_1 is called the *linear effect coefficient* and β_2 is called the *quadratic effect coefficient*.

Figure: Examples of quadratic response functions.



From Applied Linear Statistical Models by Kutner, Nachtsheim, Neter and Li

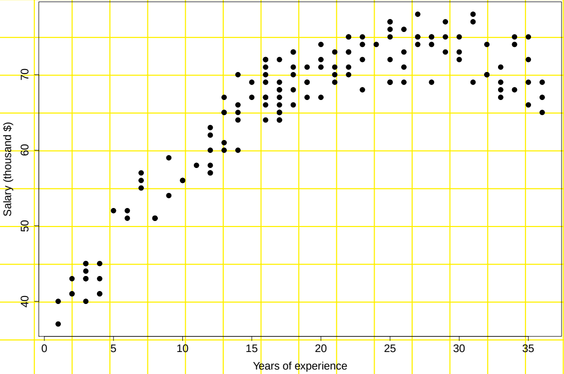
Salary

Professional organizations regularly survey their members for information concerning salaries, pensions, and conditions of employment. One goal is to relate salary to years of experience. This data has years of experience (X) and salary (Y) on 143 cases. [1]

Case	Salary($) Y	Experience(Years) X
1	71	26
2	69	19
3	73	22
...	...	...
141	67	16
142	71	20
143	69	31

[1] Source of data: Tryfos (1998): Methods for business analysis and forecasting

Figure: Scatter plot of salary versus years of experience



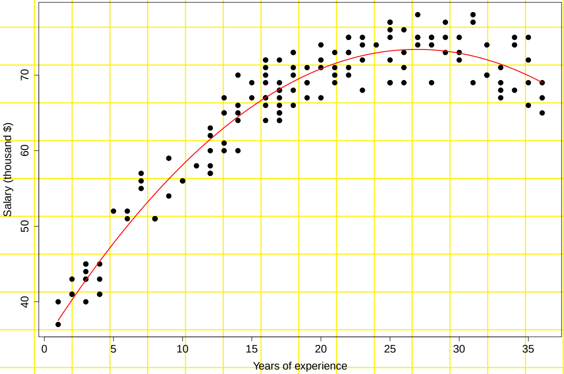
The regression relation between salary and years of experience appears to be curvilinear.

Salary: Second-Order Model

```
> salary.c=salary
> salary.c[,"Experience"]=salary[,"Experience"]-mean(salary[,"Experience"]) ## center the X variable
> fitc=lm(Salary~ Experience+I(Experience^2), data=salary.c) ## fit a second-order model
> summary(fitc)

Call:
lm(formula = Salary ~ Experience + I(Experience^2), data = salary.c)

Residuals:
Min       1Q  Median      3Q     Max
-4.5786 -2.3573  0.0957  2.0171  5.5176

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)     69.927208   0.323090  216.43   <2e-16 ***
Experience       0.861177   0.024957   34.51   <2e-16 ***
I(Experience^2) -0.053316   0.002477  -21.53   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.817 on 140 degrees of freedom
Multiple R-squared: 0.9247,     Adjusted R-squared: 0.9236
F-statistic: 859.3 on 2 and 140 DF,  p-value: < 2.2e-16
```

Figure: Fitted response function:

$$y = 69.93 + 0.861 \times (X - 18.86) - 0.0533 \times (X - 18.86)^2$$



Second-Order Model with Two Predictors

where $\tilde{X}_{i1} = X_{i1} - \overline{X}_1$, $\tilde{X}_{i2} = X_{i2} - \overline{X}_2$.

- Response function is a conic section:

$$E(Y) = \beta_0 + \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \beta_{11} \tilde{X}_1^2 + \beta_{22} \tilde{X}_2^2 + \beta_{12} \tilde{X}_1 \tilde{X}_2.$$

- This model contains separate and terms for each of the two predictors.
- It also contains a term representing the between the two predictors.
- β_{12} is called the .

Second-Order Model with Two Predictors

$$Y_i = \beta_0 + \beta_1 \tilde{X}_{i1} + \beta_2 \tilde{X}_{i2} + \beta_{11} \tilde{X}_{i1}^2 + \beta_{22} \tilde{X}_{i2}^2 + \beta_{12} \tilde{X}_{i1} \tilde{X}_{i2} + \epsilon_i, i = 1, \cdots, n,$$

where $\tilde{X}_{i1} = X_{i1} - \overline{X}_1$, $\tilde{X}_{i2} = X_{i2} - \overline{X}_2$.

- Response function is a conic section:

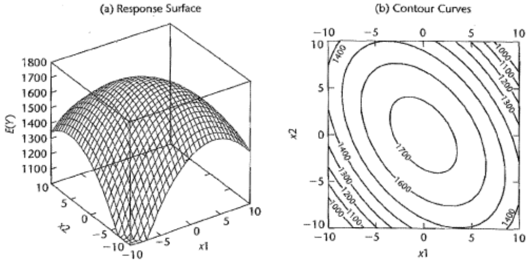
$$E(Y) = \beta_0 + \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \beta_{11} \tilde{X}_1^2 + \beta_{22} \tilde{X}_2^2 + \beta_{12} \tilde{X}_1 \tilde{X}_2.$$

- This model contains separate linear and quadratic terms for each of the two predictors.
- It also contains a cross-product term representing the interaction between the two predictors.
- β_{12} is called the *interaction effect coefficient*.

Extensions: Second-order model with K predictors; Third-order models, etc.

Figure: A quadratic response surface.

FIGURE 8.3 Example of a Quadratic Response Surface—$E\{Y\} = 1{,}740 - 4x_1^2 - 3x_2^2 - 3x_1x_2$.



From Applied Linear Statistical Models by Kutner, Nachtsheim, Neter and Li

The contour curves show various combinations of the values of the two predictors that yield the same value of the response function.

Second-Order Model with K Predictors

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k \tilde{X}_{ik} + \sum_{k=1}^{K} \beta_{kk} \tilde{X}_{ik}^2 + \sum_{1 \le k < k' \le K} \beta_{kk'} \tilde{X}_{ik} \tilde{X}_{ik'} + \epsilon_i, i = 1, \cdots, n,$$

where $\tilde{X}_{ik} = X_{ik} - \overline{X}_k$ $(k = 1, \cdots, K)$.

- Response function:

$$E(Y) = \beta_0 + \sum_{k=1}^{K} \beta_k \tilde{X}_k + \sum_{k=1}^{K} \beta_{kk} \tilde{X}_k^2 + \sum_{1 \le k < k' \le K} \beta_{kk'} \tilde{X}_k \tilde{X}_{k'}.$$

- β_ks are linear effect coefficients; β_{kk}s are quadratic effect coefficients.
- $\{\beta_{kk'} : 1 \le k < k' \le K\}$ are interaction effect coefficients between respective pairs of predictors. (The cross-product terms are second-order terms.)

Salary: Third-Order Model

```
> fit3=lm(Salary~ Experience+I(Experience^2)+I(Experience^3), data=salary.c)
> summary(fit3)
Call:
lm(formula = Salary ~ Experience + I(Experience^2) + I(Experience^3),
data = salary.c)
...
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)      69.9484745  0.3224575  216.92   <2e-16 ***
Experience        0.9364986  0.0603531   15.52   <2e-16 ***
I(Experience^2) -0.0537196  0.0024866  -21.60   <2e-16 ***
I(Experience^3) -0.0003957  0.0002888   -1.37    0.173
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.808 on 139 degrees of freedom
Multiple R-squared: 0.9257,     Adjusted R-squared: 0.9241
F-statistic: 577.1 on 3 and 139 DF,  p-value: < 2.2e-16

> anova(fit3)
Analysis of Variance Table
Response: Salary
Df Sum Sq Mean Sq   F value Pr(>F)
Experience        1 9962.9   9962.9 1263.1043 <2e-16 ***
I(Experience^2)   1 3677.9   3677.9  466.2810 <2e-16 ***
I(Experience^3)   1   14.8     14.8    1.8764  0.173
Residuals       139 1096.4      7.9
```

- First test whether the third-order term may be dropped.
  - Full model: third-order model vs. reduced model: second-order model.
  - $SSR(X^3|X, X^2) = 14.8$ with d.f. 1 and $SSE(X, X^2, X^3) = 1096.4$ with d.f. 139. The F-statistic is 1.876 and pvalue is 0.173.
  - Therefore, the third-order term is not significant and may be dropped.
- Then test whether the second-order term may be dropped.
  - Full model: second-order model vs. reduced model: first-order model.
  - $SSR(X^2|X) = 3677.9$ with d.f. 1 and $SSE(X, X^2) = SSE(X, X^2, X^3) + SSR(X^3|X, X^2) = 1111.2$ with d.f. 140. The F-statistic is 466.28 and pvalue $< 2e-16$.
  - So the second-order term is significant and should be retained.
- Thus the first-order term should also be retained and we end up with the second-order model.

Qualitative Predictors

Qualitative variables, a.k.a. **categorical variables**, represent certain characteristics of a subject.

- A qualitative variable has a fixed set of possible values/levels/classes.
- If a qualitative variable takes on exactly two values, it is called a **binary variable.**
- Examples.
    - Blood type: A, B, AB or O.
    - Smoke status: smoke or not smoke; binary variable.
    - Income level: high, medium or low.
    - Education level: high school, college, or advanced degree.

Indicators for Qualitative Variables

- To use a qualitative variable in a regression model as a predictor, we need to its classes.
- One popular approach is to use indicator variables (a.k.a. **dummy variables**).
  - An **indicator variable** is a variable only takes on the values .

Indicators for Qualitative Variables

- To use a qualitative variable in a regression model as a predictor, we need to quantitatively identify its classes.
- One popular approach is to use indicator variables (a.k.a. **dummy variables**).
  - An **indicator variable** is a variable only takes on the values 0 or 1.

To quantify a binary variable, we need indicator variable.

- Suppose the two classes are labelled as C_1, C_2. Then the indicator variable can be defined as

- For example, to code gender,

$$X = \begin{cases} 1 & \textit{if} \quad \textit{male} \\ 0 & \textit{if} \quad \textit{female} \end{cases}$$

- The above coding is , since we can arbitrarily choose the *reference class* – the class coded as 0.

To quantify a binary variable, we need **one** indicator variable.

- Suppose the two classes are labelled as C_1, C_2. Then the indicator variable can be defined as

$$X = \begin{cases} 1 & \textit{if} \quad C_1 \\ 0 & \textit{if} \quad C_2 \end{cases}$$

- For example, to code gender,

$$X = \begin{cases} 1 & \textit{if} \quad \textit{male} \\ 0 & \textit{if} \quad \textit{female} \end{cases}$$

- The above coding is **not unique**, since we can arbitrarily choose the *reference class* – the class coded as 0.

Qualitative Variables with More than Two Classes

A qualitative variable with r classes, labeled as $C_I, \cdots, C_r$, need to be represented by indicator variables, each taking on the values :

For C_r (the reference class), .

The above quantification is not unique as we can choose a different *reference class*.

Qualitative Variables with More than Two Classes

A qualitative variable with r classes, labeled as $C_1, \cdots, C_r$, need to be represented by $\mathbf{r-1}$ indicator variables, each taking on the values 0 and 1:

$$
\begin{aligned}
X_1 &= \begin{cases} 1 & if \quad C_1 \\ 0 & if \quad otherwise \end{cases} \\
X_2 &= \begin{cases} 1 & if \quad C_2 \\ 0 & if \quad otherwise \end{cases} \\
&\vdots \\
X_{r-1} &= \begin{cases} 1 & if \quad C_{r-1} \\ 0 & if \quad otherwise \end{cases}
\end{aligned}
$$

For C_r (the reference class), $X_1 = \cdots = X_{r-1} = 0$.
The above qualification is not unique as we can choose a different *reference class*.

Insurance

An economist wanted to relate the speed with which a particular insurance innovation is adopted by an insurance firm (Y) to the size of the firm (X_1) and the type of the firm (X_2). He collected data on 20 insurance firms, 10 stock firms and 10 mutual firms.

- Y – – number of months elapsed before the firm adopted the innovation and X_1 – – the amount of total assets of the firm are quantitative variables.
- Type of the firm is a variable taking on two values: "stock" or "mutual". If we choose "mutual" as the reference class, then it can be quantified by an indicator variable:

Insurance

An economist wanted to relate the speed with which a particular insurance innovation is adopted by an insurance firm (Y) to the size of the firm (X_1) and the type of the firm (X_2). He collected data on 20 insurance firms, 10 stock firms and 10 mutual firms.

- Y – – number of months elapsed before the firm adopted the innovation and X_1 – – the amount of total assets of the firm are quantitative variables.
- Type of the firm is a binary variable taking on two values: "stock" or "mutual". If we choose "mutual" as the reference class, then it can be quantified by an indicator variable:

$$X_2 = \begin{cases} 1 & \textit{if} \quad \textit{stock} \\ 0 & \textit{if} \quad \textit{mutual} \end{cases}$$

A snapshot of the data.

Firm	Number_of_month_elapsed	Firm_size	Firm_Type	Indicator_Code
	Y	X1		X2
1	17	151	mutual	0
2	26	92	mutual	0
3	21	175	mutual	0
..	...	...	...	...
18	13	305	stock	1
19	30	124	stock	1
20	14	246	stock	1

Figure: Boxplots

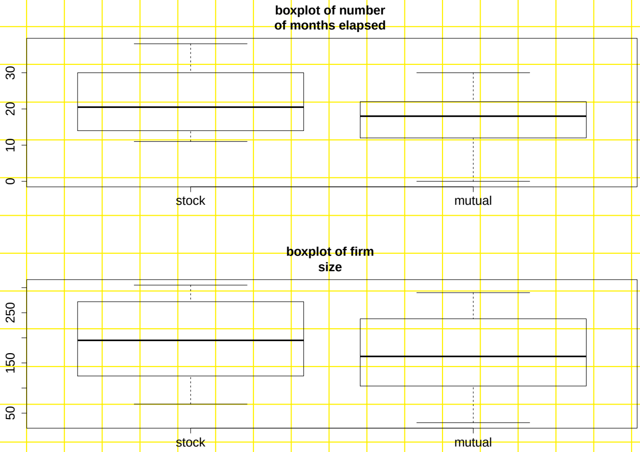
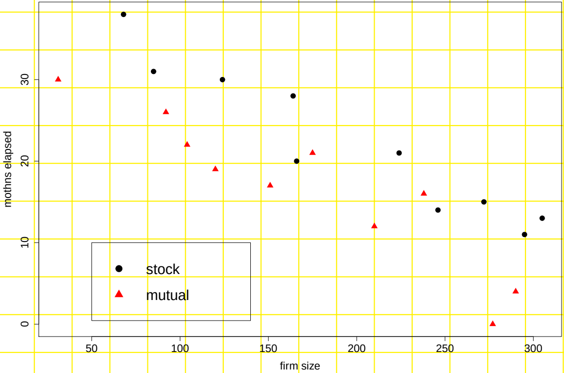
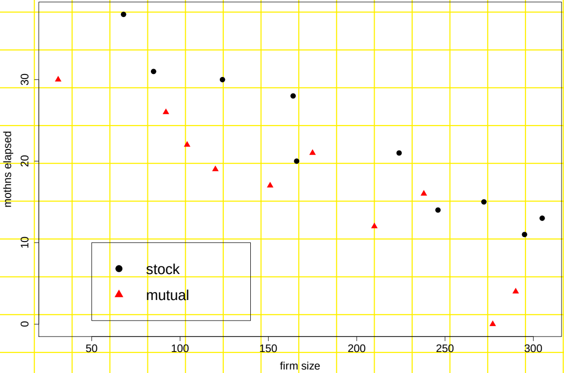


Figure: Scatter plot of months elapsed (Y) versus firm size (X_1).



The slope appears to be for the two types of firms, whereas the intercept appears to be . This means that a model would suffice.

Figure: Scatter plot of months elapsed (Y) versus firm size (X_1).



The slope appears to be similar for the two types of firms, whereas the intercept appears to be different. This means that a first-order model would suffice.

A first order model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- The response function (mean response):
  - For mutual firms, $X_2 = 0$ and the response function becomes
  - For stock firms, $X_2 = 1$ and the response function becomes
  - Both are with

A first order model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- The response function (mean response):

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

  - For mutual firms, $X_2 = 0$ and the response function becomes

$$E(Y) = \beta_0 + \beta_1 X_1 \quad \text{mutual firms}$$

  - For stock firms, $X_2 = 1$ and the response function becomes

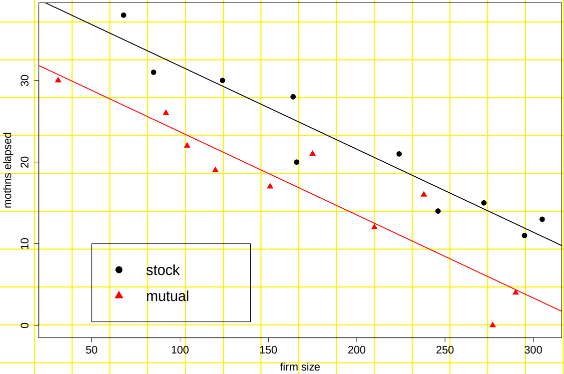
$$E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{stock firms}$$

  - Both are straight lines with the same slope β_1 but with intercepts differing by β_2.

Insurance: First Order Model

```
> data=data.frame(read.table("insurance.txt"))
> names(data)=c("Y", "X1", "X2")
> fit=lm(Y~ X1+factor(X2), data=data)
> summary(fit)

Call:
lm(formula = Y ~ X1 + factor(X2), data = 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 ***
X1              -0.101742   0.008891 -11.443 2.07e-09 ***
factor(X2)stock  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
```

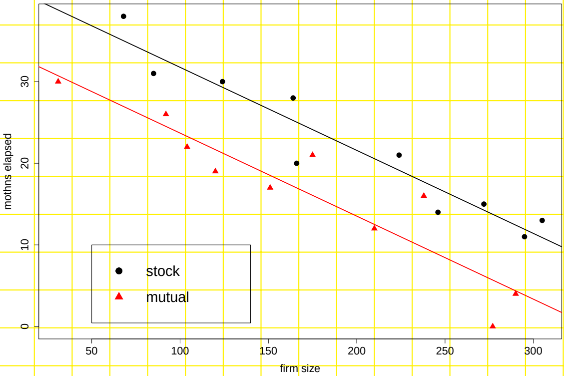
Figure: Response functions for the stock firms (black) and mutual firms (red).



Stock firms response line: .

Mutual firms response line: .

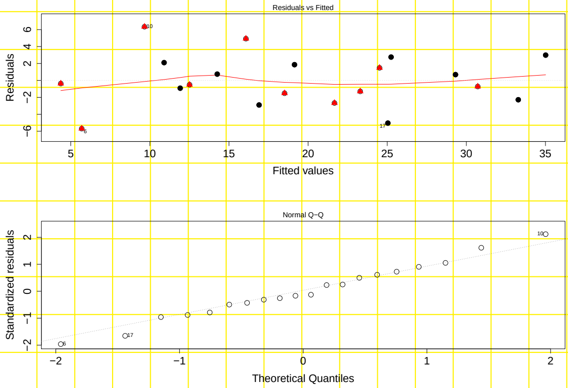
Figure: Response functions for the stock firms (black) and mutual firms (red).



Stock firms response line: $\widehat{E(Y)} = (33.874 + 8.055) - 0.1017X_1$.

Mutual firms response line: $\widehat{E(Y)} = 33.874 - 0.1017X_1$.

Figure: Residual plots.



No obvious violation of the linearity, constant error variance and normal error assumptions.

The economist was most interested in the effect of firm type on the speed to adopt an innovation.

- $\hat{\beta}_2 = 8.055$ means that for any given firm size, on average, it takes stock firms to adopt an innovation than mutual firms of the same size.
- A 95% confidence interval for β_2: $t(0.975; 17) = 2.11$

  With 95% confidence, we conclude that on average stock firms takes to adopt an innovation than mutual firms.
- The pvalue for testing whether $\beta_2 = 0$ is 3.74×10^{-5}. Therefore, β_2 is highly significant and firm type has a effect on the speed of adopting an innovation.

Why not simply fit two separate regression models for stock firms and mutual firms?

The economist was most interested in the effect of firm type on the speed to adopt an innovation.

- $\hat{\beta}_2 = 8.055$ means that for any given firm size, on average, it takes stock firms 8 more months to adopt an innovation than mutual firms of the same size.
- A 95% confidence interval for β_2: $t(0.975; 17) = 2.11$

$$8.055 \pm 2.11 \times 1.459 = [4.98, 11.13].$$

  With 95% confidence, we conclude that on average stock firms takes between 5 to 11 more months to adopt an innovation than mutual firms.
- The pvalue for testing whether $\beta_2 = 0$ is 3.74×10^{-5}. Therefore, β_2 is highly significant and firm type has a significant effect on the speed of adopting an innovation.

Why not simply fit two separate regression models for stock firms and mutual firms?

Summary

Interpretation of regression coefficients in a first order model with a quantitative variable (X_1) and an indicator variable (X_2):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, n.$$

- β_1 is the common slope of the mean response line under both classes.
- β_0 is the baseline intercept under class 0 (i.e., the reference class).
- β_2 shows how much higher (if positive) or lower (if negative) the mean response line is for class 1 for any given value of X_1.
- The effect of one variable is **the same** no matter the value of the other variable.

Interactions between Quantitative and Qualitative Predictors

Interaction between qualitative and quantitative predictors can be introduced into the model through the usual manner, by

.

- Insurance company. A model with interaction between firm size and firm type:

  where X_1 is the amount of assets of a firm and X_2 is an indicator variable indicating the type of a firm.

Interactions between Quantitative and Qualitative Predictors

Interaction between qualitative and quantitative predictors can be introduced into the model through the usual manner, by adding cross-product terms.

- Insurance company. A model with interaction between firm size and firm type:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20,$$

  where X_1 is the amount of assets of a firm and X_2 is an indicator variable indicating the type of a firm.

- Response function (mean response):

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2.$$

- For mutual firms, $X_2 = 0$ and thus $X_1 X_2 = 0$. The response function becomes

  which is a straight line with slope β_1 and intercept β_0.
- For stock firms, $X_2 = 1$ and thus $X_1 X_2 = X_1$. The response function becomes

  which is a straight line with slope $\beta_1 + \beta_3$ and intercept $\beta_0 + \beta_2$.

- Response function (mean response):

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2.$$

- For mutual firms, $X_2 = 0$ and thus $X_1 X_2 = 0$. The response function becomes

$$E(Y) = \beta_0 + \beta_1 X_1 \quad \text{mutual firms,}$$

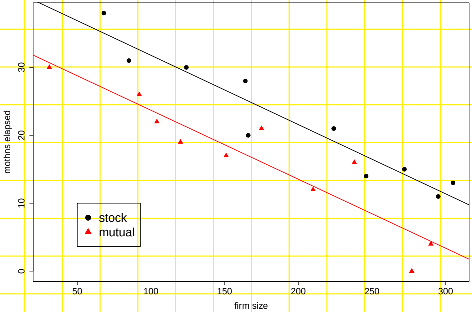
which is a straight line with slope β_1 and intercept β_0.

- For stock firms, $X_2 = 1$ and thus $X_1 X_2 = X_1$. The response function becomes

$$E(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 \quad \text{stock firms,}$$

which is a straight line with slope $\beta_1 + \beta_3$ and intercept $\beta_0 + \beta_2$.

Figure: Response functions for the stock firms (black) and mutual firms (red) under the interaction model.



Stock firms: .

Mutual firms:

The two lines are because

is very small compared to .

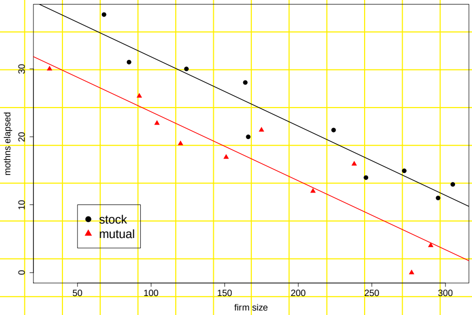
Insurance: Interaction Model

```
> fit2=lm(Y~ X1+factor(X2)+X1:factor(X2), data=data)
> summary(fit2)
Call:
lm(formula = Y ~ X1 + factor(X2) + X1:factor(X2), data = data)

Residuals:
Min       1Q  Median      3Q     Max
-5.7144 -1.7064 -0.4557  1.9311  6.3259

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)         33.8383695  2.4406498  13.864 2.47e-10 ***
X1                  -0.1015306  0.0130525  -7.779 7.97e-07 ***
factor(X2)stock      8.1312501  3.6540517   2.225   0.0408 *
X1:factor(X2)stock -0.0004171  0.0183312  -0.023   0.9821
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
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
```

β_3 and we conclude that there is

and the first-order model .

Figure: Response functions for the stock firms (black) and mutual firms (red) under the interaction model.



Stock firms: $\widehat{E(Y)} = (33.838 + 8.131) - (0.1015 + 0.00042)X_1$.
Mutual firms: $\widehat{E(Y)} = 33.838 - 0.1015X_1$. The two lines are nearly parallel because $\hat{\beta}_3 = -0.00042$ is very small compared to $\hat{\beta}_1 = -0.1015$.

Insurance: Interaction Model

```
> fit2=lm(Y~ X1+factor(X2)+X1:factor(X2), data=data)
> summary(fit2)
Call:
lm(formula = Y ~ X1 + factor(X2) + X1:factor(X2), data = data)

Residuals:
Min      1Q  Median      3Q     Max
-5.7144 -1.7064 -0.4557  1.9311  6.3259

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)        33.8383695  2.4406498  13.864 2.47e-10 ***
X1                 -0.1015306  0.0130525  -7.779 7.97e-07 ***
factor(X2)stock     8.1312501  3.6540517   2.225   0.0408 *
X1:factor(X2)stock -0.0004171  0.0183312  -0.023   0.9821
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
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
```

β_3 is not significant and we conclude that there is no interaction effect and the first-order model suffices.

Summary

Interpretation of regression coefficients in an interaction model with a quantitative variable (X_1) and an indicator variable (X_2).:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, n$$

- β_0 and β_1 are baseline intercept and slope, respectively, of the response function for class 0 (i.e., the reference class).
- β_2 indicates how much greater (if positive) or smaller (if negative) is the intercept of the response function for class 1.
- β_3 indicates how much greater (if positive) or smaller (if negative) is the slope of the response function for class 1.
- The effect of one variable depends on the value of the other variable.