Chapter 8 Regression Models for Quantitative and Qualitative Predictors

Topics:

- Use of interaction and polynomial terms for quantitative predictors
- Use of indicator variables for qualitative predictors
- Regression models containing both quantitative and qualitative predictors

8.1 Polynomial Regression

- Polynomial Models are used when the true linear regression function is a polynomial or the true regression function can be approximated by a polynomial

- We can fit a quadratic, cubic, etc. relationship by defining squares, cubes, etc., and use them as additional explanatory variables

- We can also do this with more than one explanatory variable, in which case we also often include an interaction term. When we do this we generally create a multicollinearity problem, which can often be corrected by standardization or centering.

Figure 1: Simulated data study

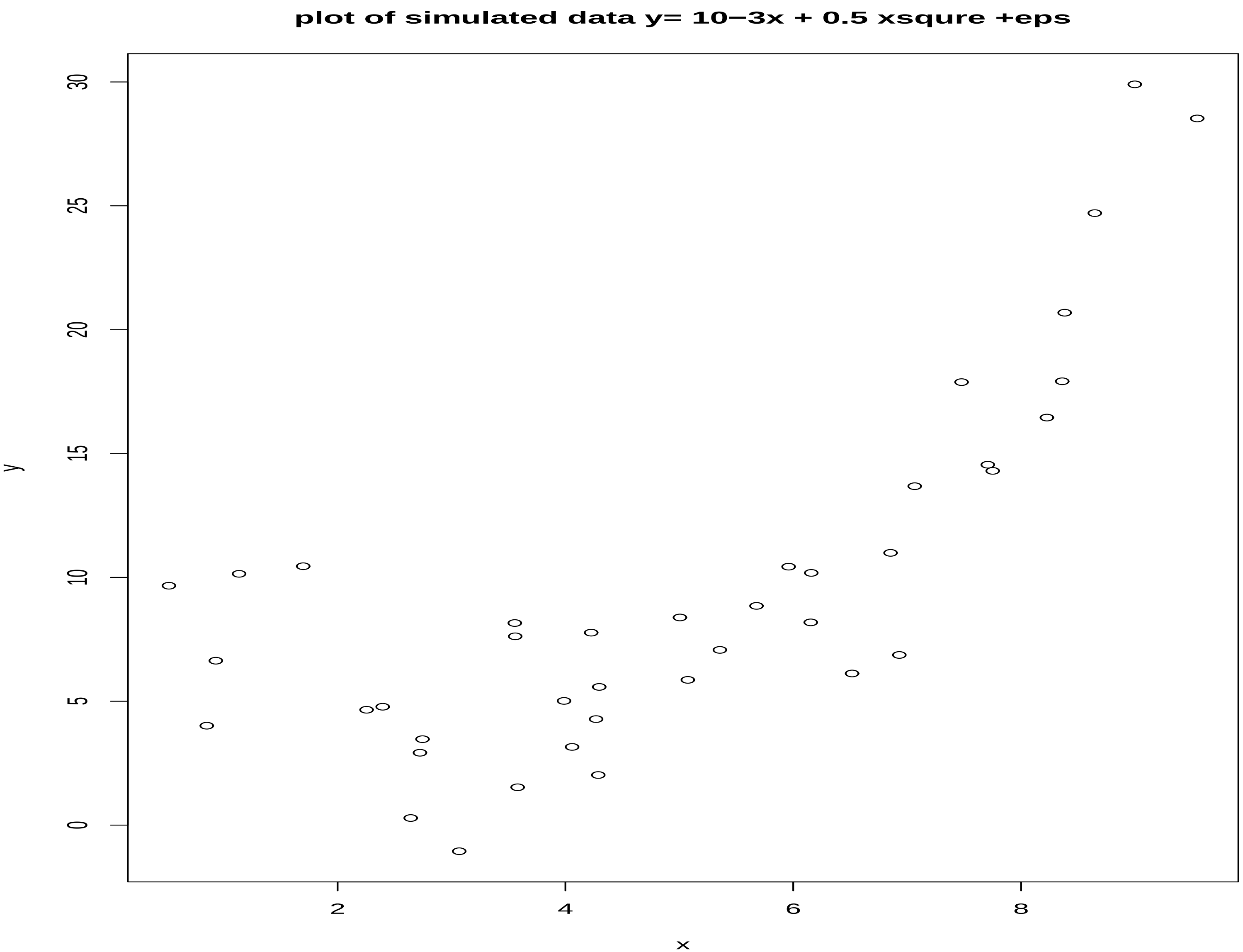
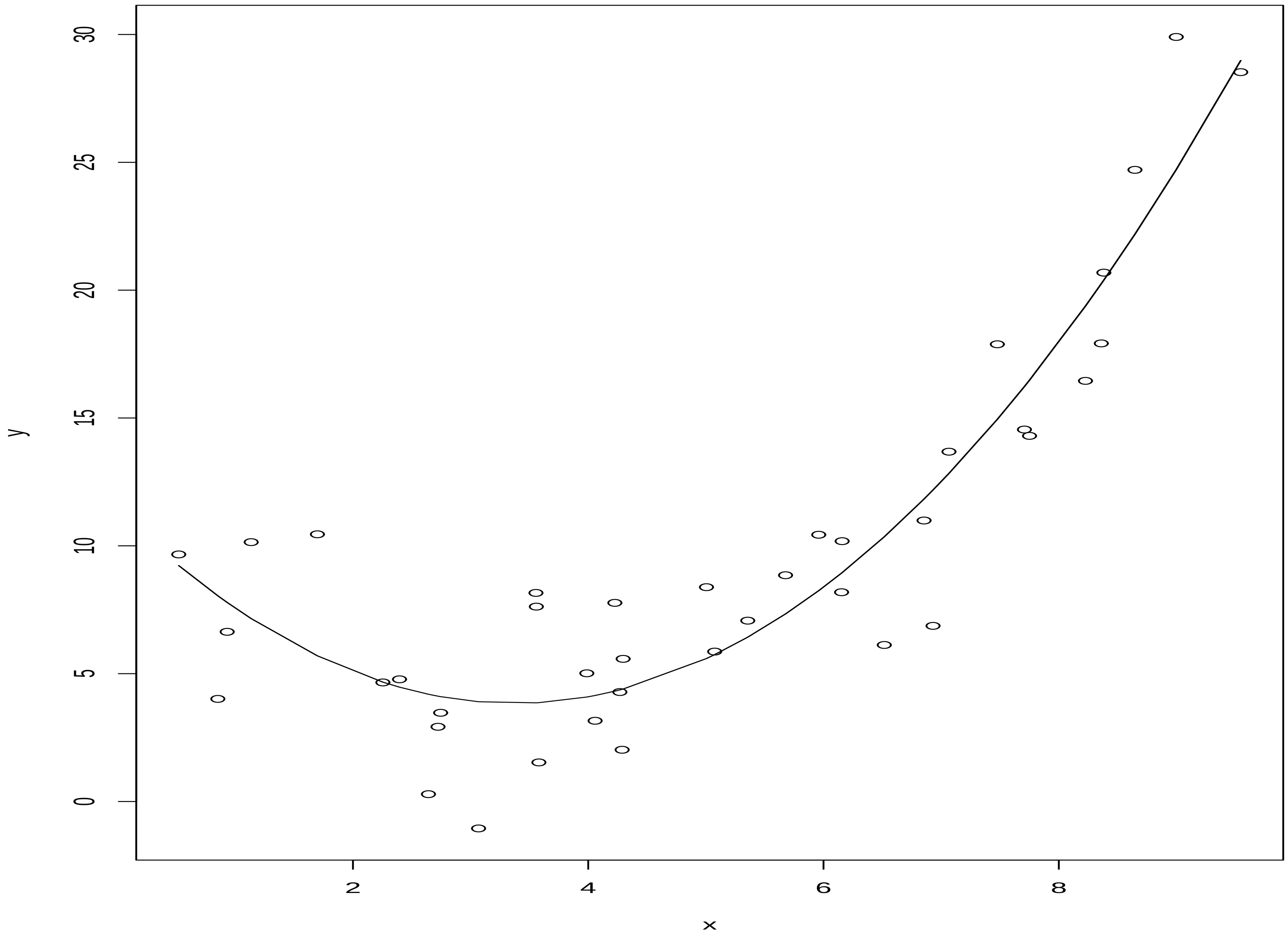


Figure 2: Simulated data study 2



Note that

- a main danger in using polynomial regression models, is that extrapolations may be hazardous with these models, especially those with higher-order terms.

- polynomial regression models may provide good fits for the data at hand, but may turn in unexpected directions when extrapolated beyond the range of the data.

Example 7.3.1. Hooker data (Christensen 1st edition)

Forbes (1857) reported data on the relationship between atmospheric pressure and the boiling point of water that were collected in the Himalaya mountains by Joseph Hooker. Weisberg (1985, p. 28) presented a subset of 31 observations that are used as our example.

Figure 3: Hooker data with fitted regression line

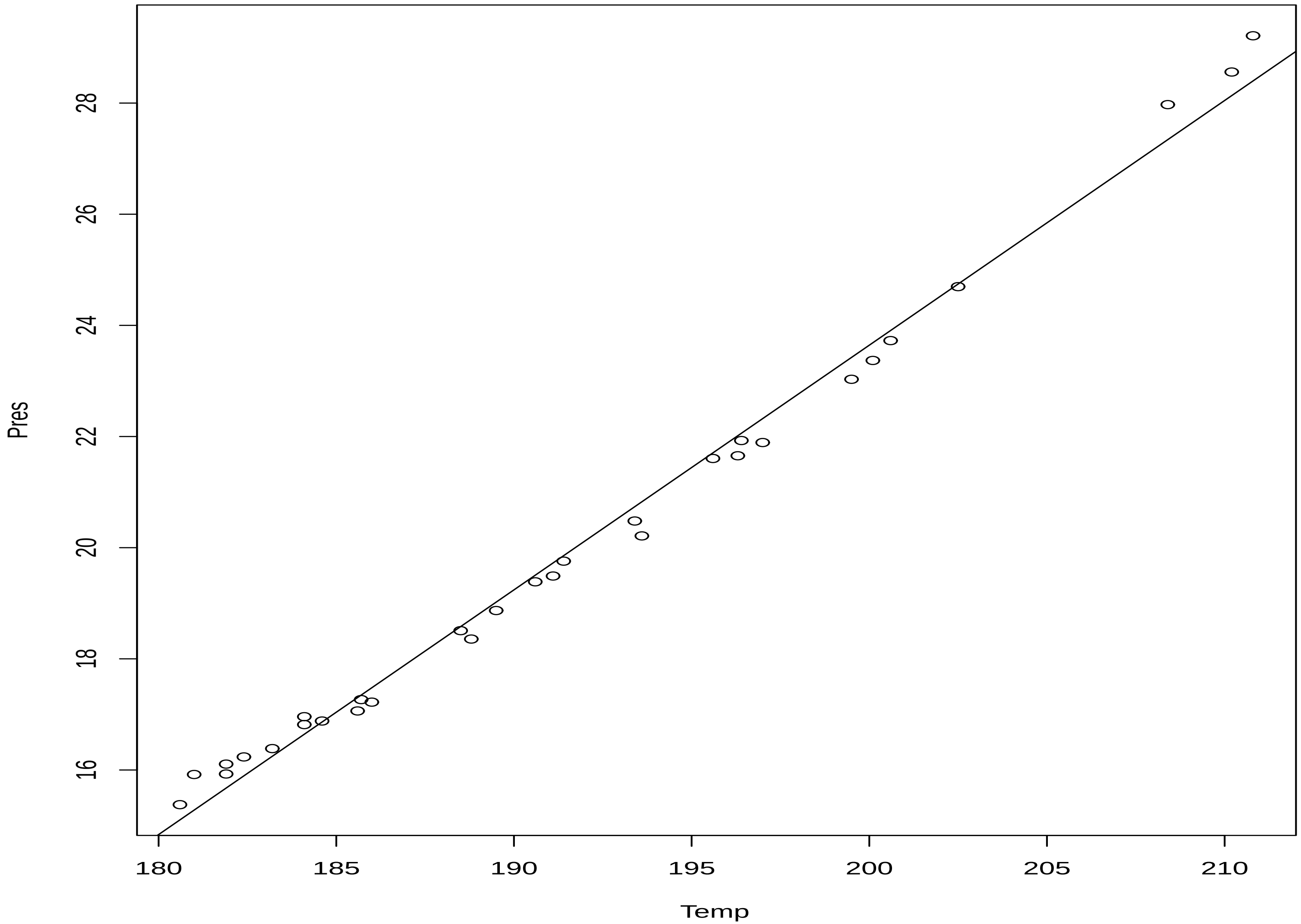
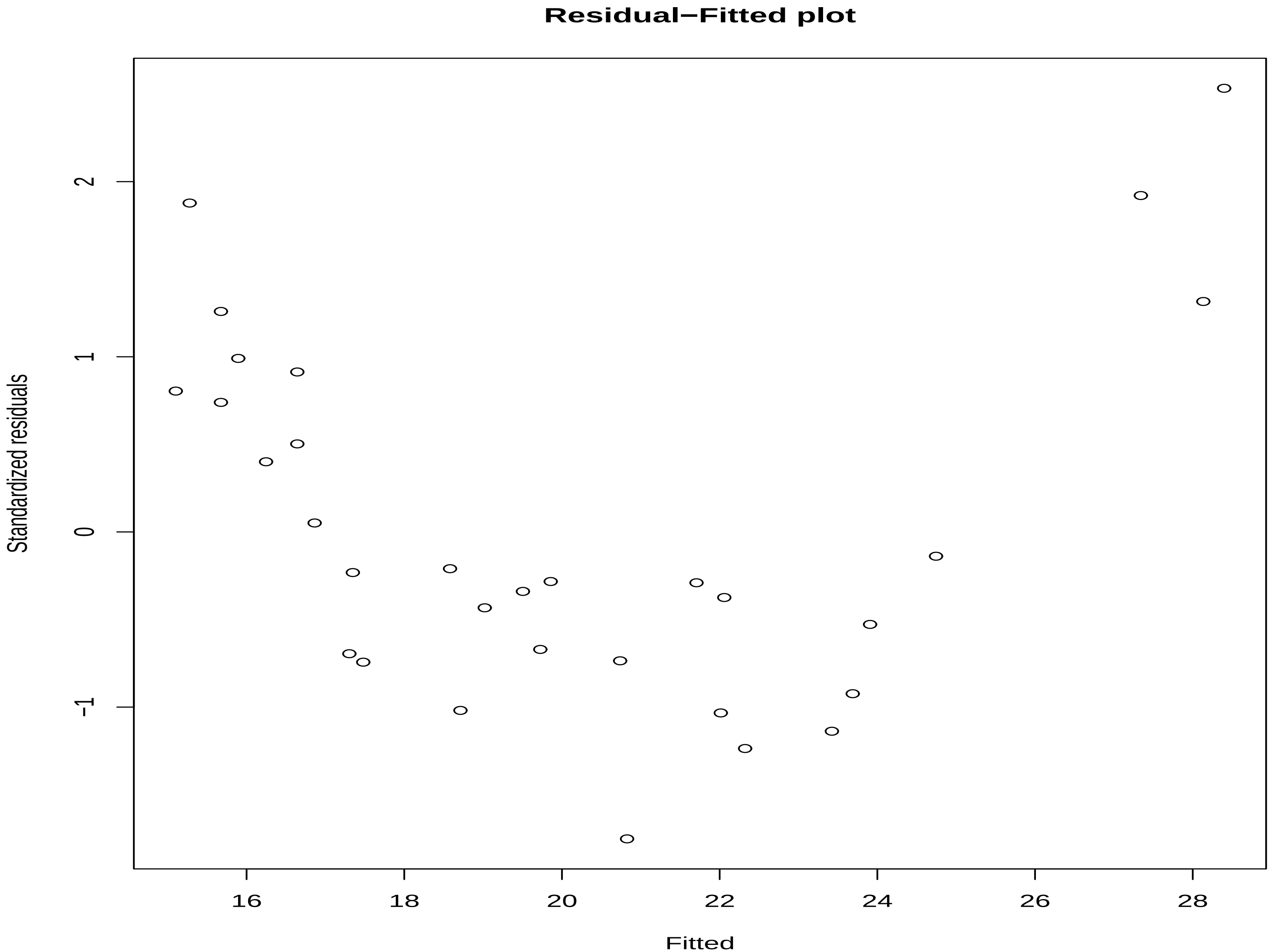


Figure 4: Hooker data: residual vs fitted values



Note that

- The simple linear regression of pressure on temperature shows a lack of fit.

- The residual plot shows nonrandom structure.

- Try to eliminate lack of fit in the simple linear regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ by fitting a larger model.

- Try quadratic model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$
  or cubic model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i$
  or higher degree polynomials.

Fit a fifth degree (quintic) polynomial to Hooker's data.

$$y_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 x_i^3 + \gamma_4 x_i^4 + \gamma_5 x_i^5 + \epsilon_i \tag{1}$$

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \beta_2(x_i - \bar{x})^2 + \beta_3(x_i - \bar{x})^3 + \beta_4(x_i - \bar{x})^4$$
$$+\beta_5(x_i - \bar{x})^5 + \epsilon_i \tag{2}$$

Centering:

- Remove or reduce the correlation between explanatory variables and their interactions by centering: subtracting the mean from x_i, $x_i - \bar{x}$.
- Model (1) and Model (2) are equivalent in that they always give the same fitted values, residuals and degrees of freedom.

Table 1: Table of coefficient: Hooker data, quintic model

Predictor	$\hat{\beta}_k$	$SE(\hat{\beta}_k)$	t	P
Constant	-59.911	2.337	-25.63	0.000
$(x - \bar{x})$	0.41540	0.01216	34.17	0.000
$(x - \bar{x})^2$	0.002179	0.002260	0.96	0.344
$(x - \bar{x})^3$	0.0000942	0.0001950	0.48	0.633
$(x - \bar{x})^4$	0.00001523	0.00001686	0.90	0.375
$(x - \bar{x})^5$	-0.00000080	0.00000095	-0.84	0.409

Table 2: Analysis of variance: Hooker data SLR

Source	df	SS	MS	F	P
Regression	1	444.17	444.17	3497.89	0.000
Error	29	3.68	0.13		
Total	30	447.85			

Table 3: Analysis of variance: Hooker data, quintic model

Source	df	SS	MS	F	P
Regression	5	447.175	89.435	3315.48	0.000
Error	25	0.674	0.027		
Total	30	447.850			

Testing the quintic model against the original simple linear regression model or testing if $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. Recall that

$$MSTest = \frac{SSE(R) - SSE(F)}{dfE(R) - dfE(F)}$$

and

$$F = \frac{MSTest}{MSE(F)}$$

Reject H_0, if $F > F(1 - \alpha; dfE(R) - dfE(F), dfE(F))$

The F statistic is

$$\begin{aligned}
F &= \frac{\{SSE(SRL) - SSE(POL)\}/(dfE(SRL) - dfE(POL))}{MSE(POL)} \\
&= \frac{(3.68 - 0.674)/(29 - 25)}{0.027} \\
&= 27.83
\end{aligned}$$

which is much bigger $F(4, 25, 0.99) = 4.18$. The test suggests rejecting H_0.

Picking a polynomial

- The table of coefficients for the quintic polynomial on the Hooker data provides a t test for whether we can drop each variable out of the model.

- In a quintic model, the only t statistic of interest is the one that tests whether you can drop x^5 so that you could get by with a quantic polynomial.

- If you are satisfied with a quartic polynomial, it makes sense to test whether you can get by with a cubic.

- We would like to fit the sequence of models

$$y_i = \beta_0 + \epsilon_i \tag{3}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{4}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \tag{5}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i \tag{6}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \epsilon_i \tag{7}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \epsilon_i \tag{8}$$

and find the smallest model that fits the data.

Table 4: Extra sums of Model 3-Model 8

Source		Model comparison	df	Seq SS	F
$h1 = (x - \bar{x})$	$SSR(h1)$	SSE(3) -SSE(4)	1	444.167	16450.7
$h2 = (x - \bar{x})^2$	$SSR(h2 \mid h1)$	SSE(4) -SSE(5)	1	2.986	110.6
$h3 = (x - \bar{x})^3$	$SSR(h3 \mid h2, h1)$	SSE(5) -SSE(6)	1	0.000	0.0
$h4 = (x - \bar{x})^4$	$SSR(h4 \mid h3, h2, h1)$	SSE(6) -SSE(7)	1	0.003	0.1
$h5 = (x - \bar{x})^5$	$SSR(h5 \mid h4, h3, h2, h1)$	SSE(7) -SSE(8)	1	0.019	0.7

F statistic for dropping the fifth degree term from the polynomial is

$$F = \frac{SSE(7) - SSE(8)}{MSE(8)} = \frac{0.019}{0.027} = 0.7 = (-0.84)^2$$

F statistic for dropping the fourth degree term is

$$F = \frac{SSE(6) - SSE(7)}{MSE(8)} = \frac{0.003}{0.027} = 0.01$$

F statistic for dropping the quadratic term is

$$F = \frac{SSE(4) - SSE(5)}{MSE(8)} = \frac{2.986}{0.027} = 110.5926$$

Compared to $F(1, 25, 0.95) = 4.24$, we can clearly drop any of the terms down to the quadratic term.

Exploring the selected model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i.$$

Table 5: Table of coefficient: Hooker data, quadratic model.

Predictor	$\hat{\beta}_k$	$SE(\hat{\beta}_k)$	t	P
Constant	88.02	13.93	6.32	0.000
x	-1.1295	0.1434	-7.88	0.000
x^2	0.0040330	0.0003682	10.95	0.000

- t statistic for β_2 is 10.95 with $P -$ value $= 0.000$, highly significant, so the quadratic model accounts for a significant amount of the lack of fit displayed by the simple linear regression model.

Figure 5: Hooker data: scatterplot with SRL fit (blue) and quadratic fit

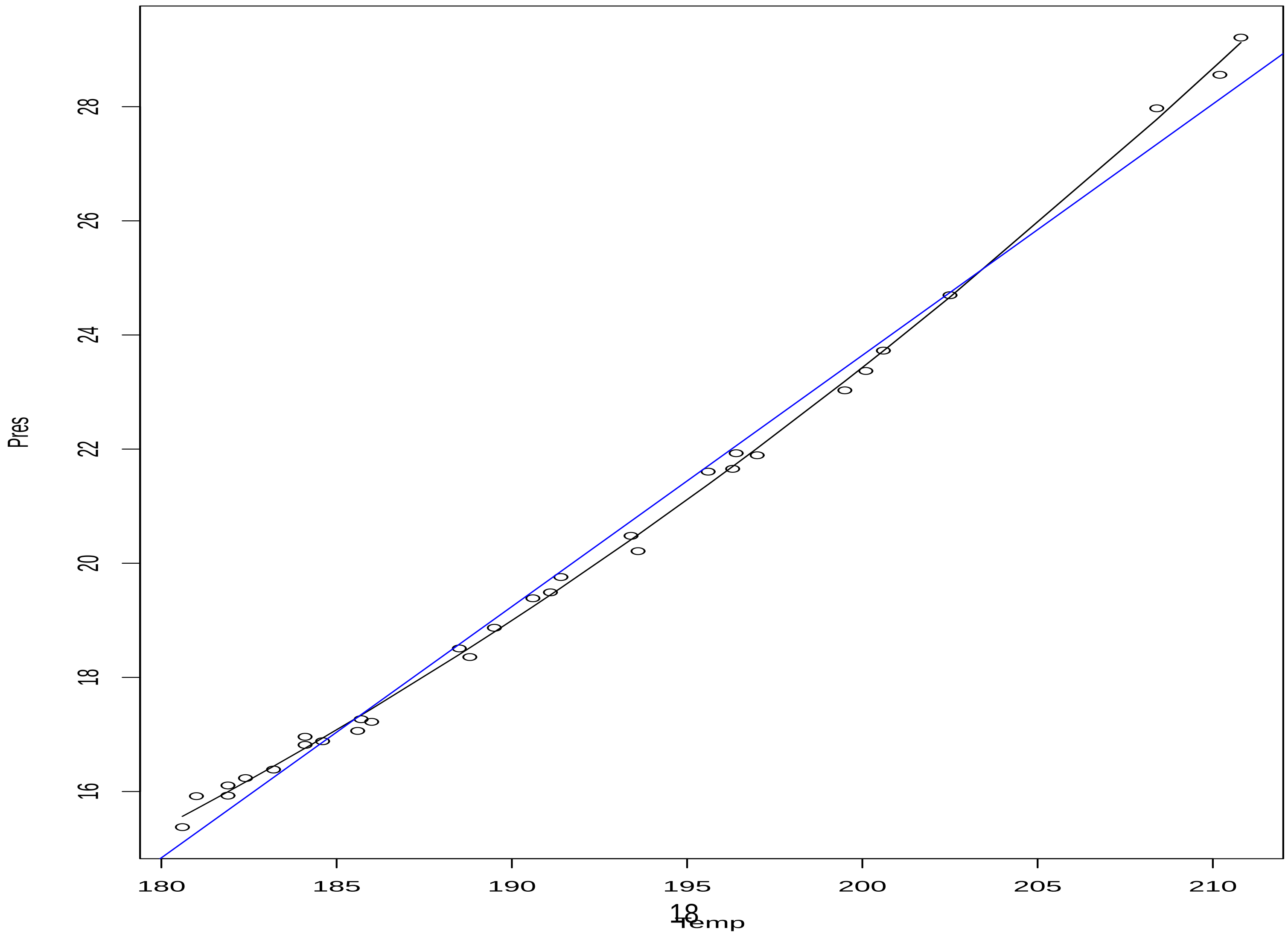


Table 6: Analysis of variance: Hooker data, quadratic model.

Source	df	SS	MS	F	P
Regression	2	447.15	223.58	8984.23	0.000
Error	28	0.70	0.02		
Total	30	447.85			

- F statistic provides a test of whether the quadratic model explains the data better than the model with only an intercept. $H_0 : \beta_1 = \beta_2 = 0$.
- Regression function $\hat{y} = 88.02 - 1.1295x + 0.004033x^2$
- Residuals: $\hat{\epsilon}_i = y_i - \hat{y}_i$
- $R^2 = SSR/SSTO = 447.15/447.85 = 99.8\%$

Figure 6: Hooker data: residual vs. fitted value, quadratic model

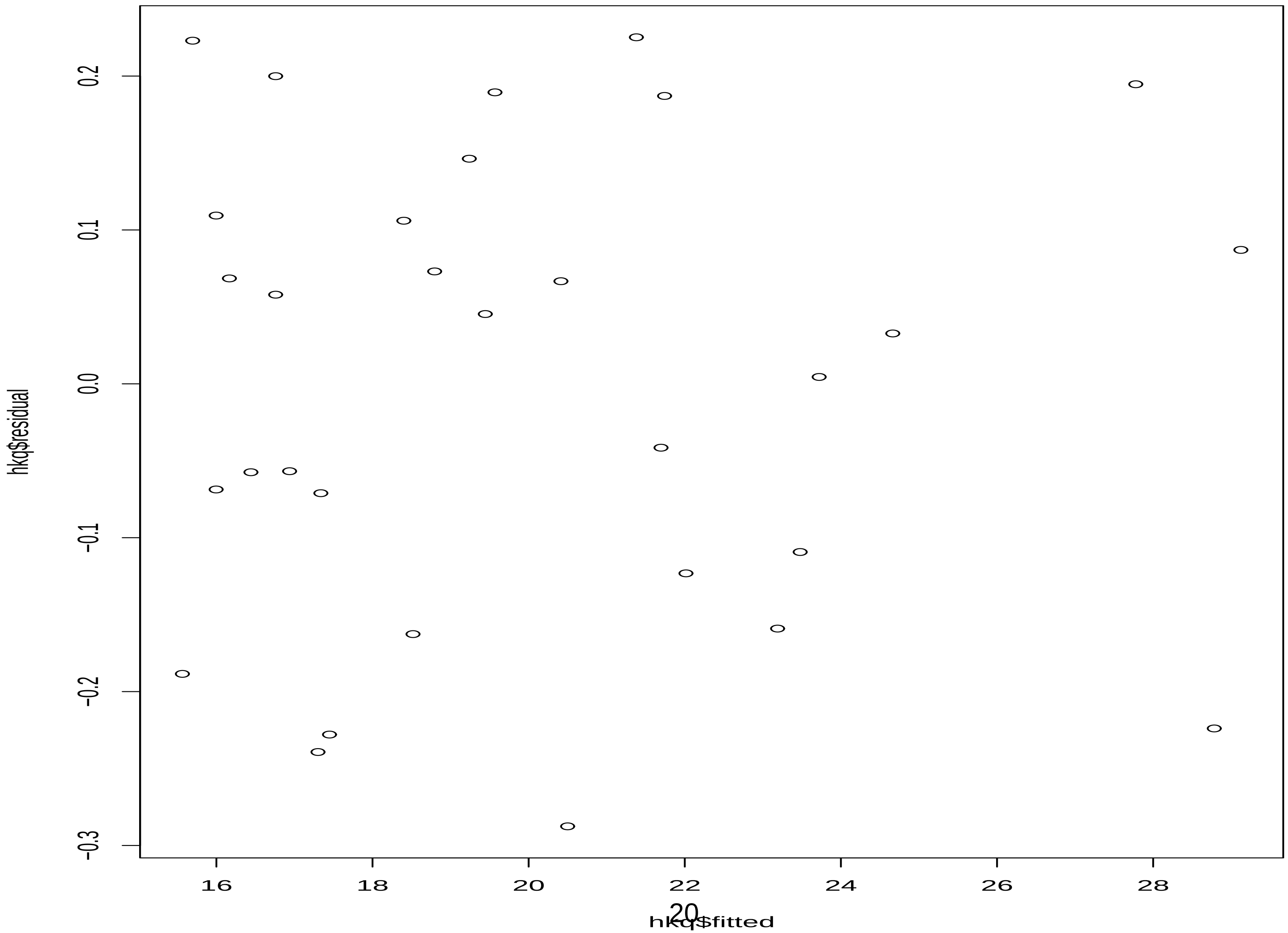
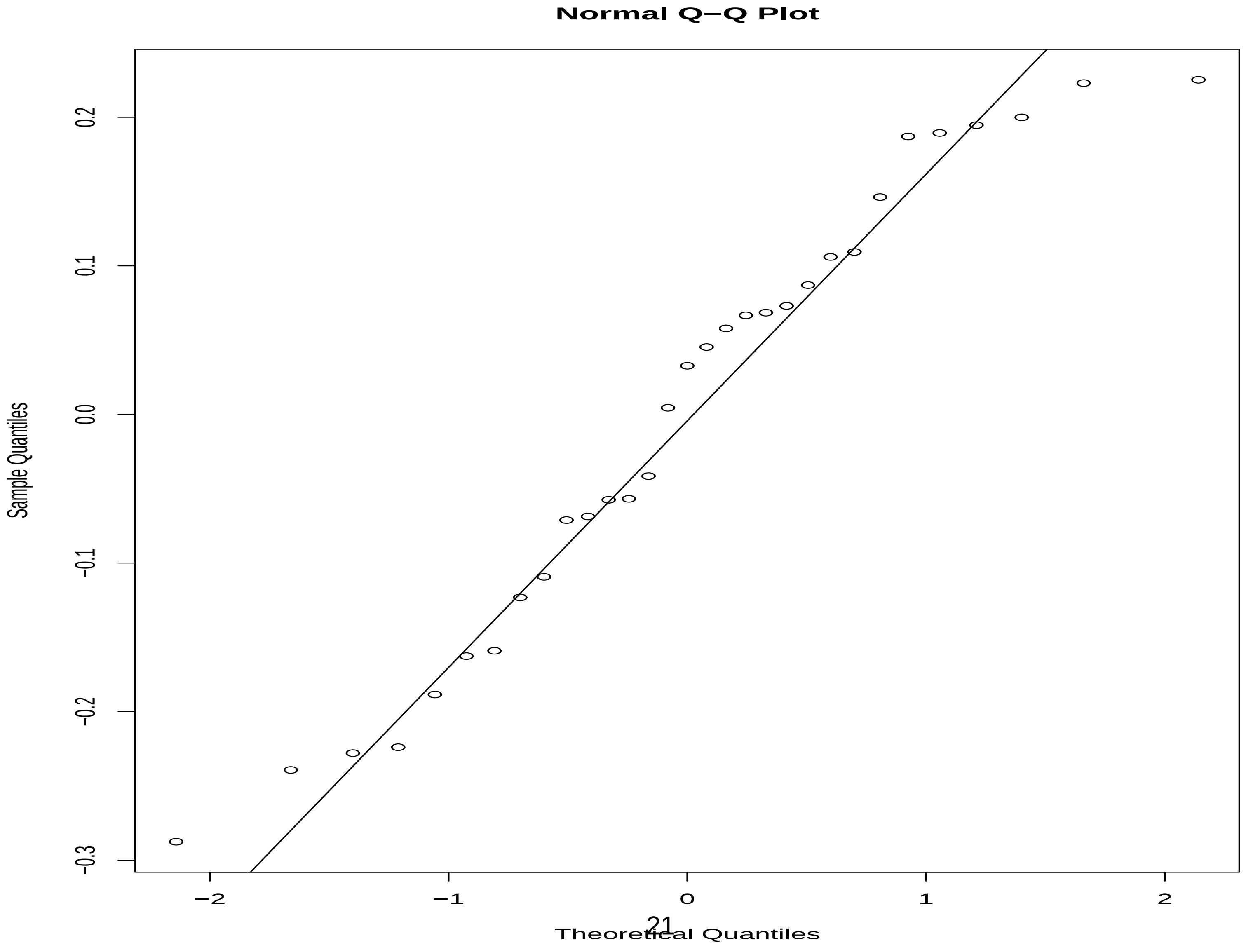


Figure 7: Hooker data: qq plot, quadratic model



- Residual against the predicted values look good

- Normal plot has a shoulder at the top but ok

- Consider testing the quadratic model for lack of fit by comparing it to the quintic model, The F statistic is

$$F = \frac{(0.70 - 0.674)/(28 - 25)}{0.027} = 0.321,$$

which is much smaller than $F(0.95, 3, 25) = 2.99$ and makes no suggestion of lack of fit.

8.2 Interaction Regression Models

Additive Model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

Interaction Model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$

- When an interaction is present, the effect of the change in the mean response when the value of a predictor variable changes depends on the value of another predictor variable.

- β_1 and β_2 no longer indicate the change in the mean response with a unit increase of the predictor variable, with the other predictor variable held constant at any given level. For example, change in the mean response with a unit increase in x_1 when x_2 is held constant is $\beta_1 + \beta_3 x_2$.

Example:

——No interaction

$$E\{y\} = 10 + 2x_1 + 5x_2$$

- when $x_2 = 1$, the response function $E\{y\}$ as a function of x_1 is

$$E\{y\} = 15 + 2x_1 \tag{9}$$

- when $x_2 = 3$,

$$E\{y\} = 25 + 2x_1 \tag{10}$$

- Line (9) and (10) are parallel.

—–Reinforcement Interaction Effect

$$E\{y\} = 10 + 2x_1 + 5x_2 + .5x_1x_2$$

- when $x_2 = 1$, the response function $E\{y\}$ as a function of x_1 is

$$E\{y\} = 15 + 2.5x_1 \tag{11}$$

- when $x_2 = 3$,

$$E\{y\} = 25 + 3.5x_1 \tag{12}$$

- Line (11) and (12) are not parallel.

——Interference Interaction Effect

$$E\{y\} = 10 + 2x_1 + 5x_2 - .5x_1x_2$$

- when $x_2 = 1$, the response function $E\{y\}$ as a function of x_1 is

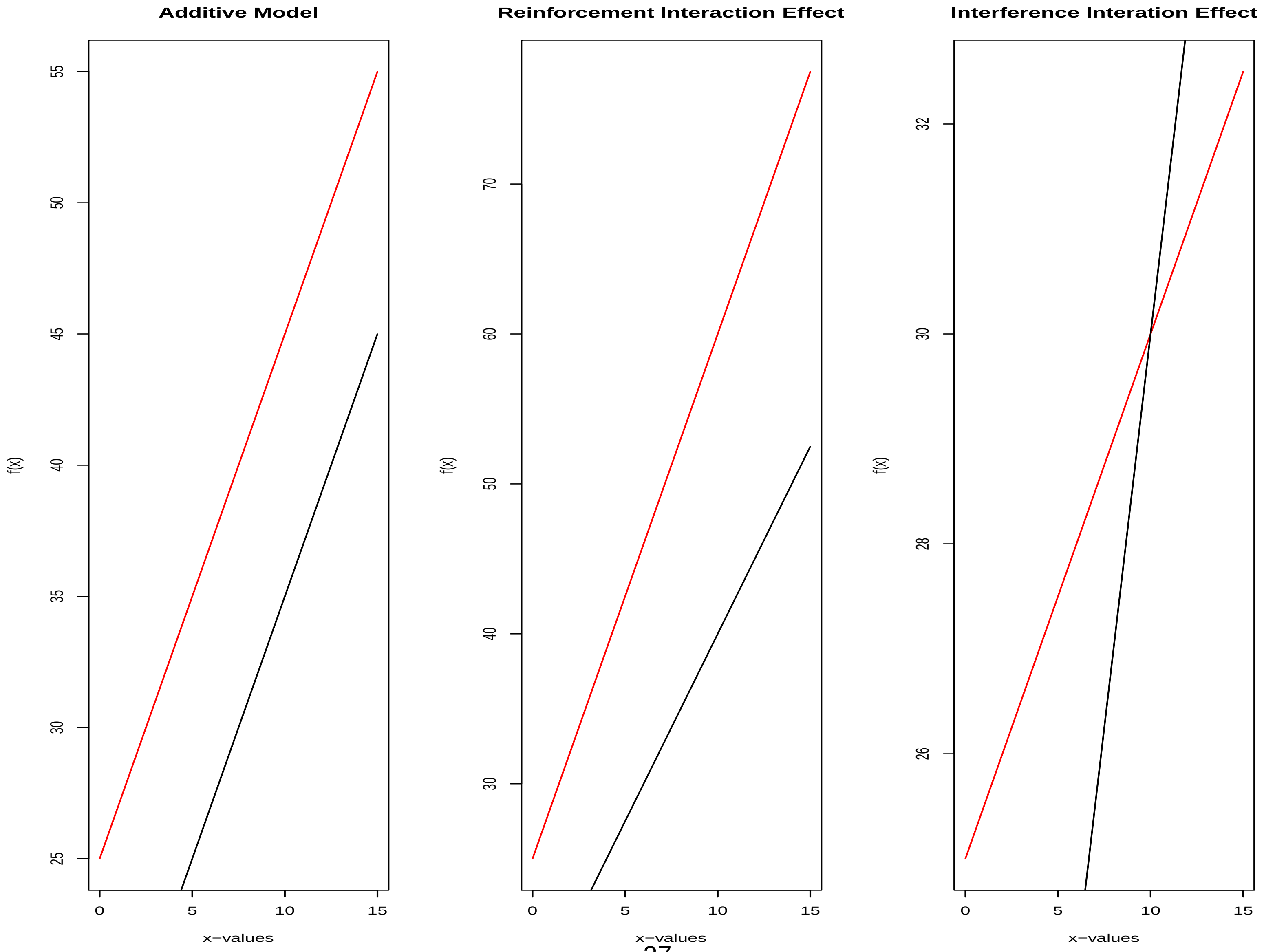
$$E\{y\} = 15 + 1.5x_1 \tag{13}$$

- when $x_2 = 3$,

$$E\{y\} = 25 + .5x_1 \tag{14}$$

- Line (13) and (14) are not parallel.

Figure 8:



Additive Model
Reinforcement Interaction Effect
Interference Interation Effect
f(x)
x−values

Body fat Example

We wish to test formally in the body fat example, whether interaction terms between the three predictor variables should be included in the regression model. Consider the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1} x_{i2}$$

$$+\beta_5 x_{i1} x_{i3} + \beta_6 x_{i2} x_{i3} + \epsilon_i$$

- The overall F test is significant, but none of the individual t tests are significant, indicating multicollinearity

- Notice some of the predictor variables are highly correlated with some of the interaction terms, and there are high correlations among the interaction terms. For example

$$r_{x1,x1x2} = .989, r_{x1x3,x2x3} = .998$$

- Check if interaction terms need to be included in the model

Figure 9: Bodyfat: residual from the additive model vs. interaction x1x2, didn't see a pattern, don't include

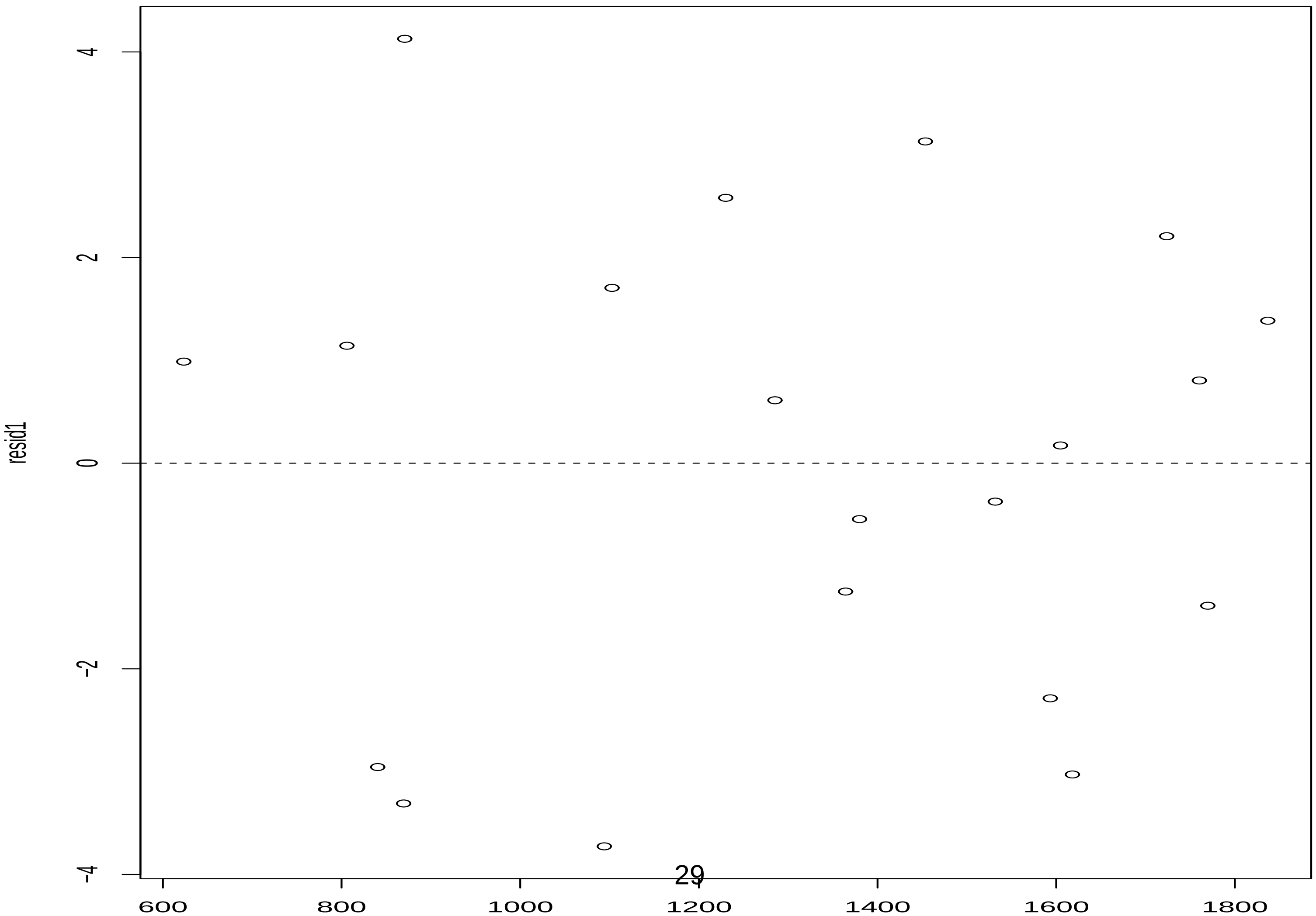


Figure 10: Bodyfat: residual from the additive model vs. interaction x1x3, didn't see a pattern, don't include

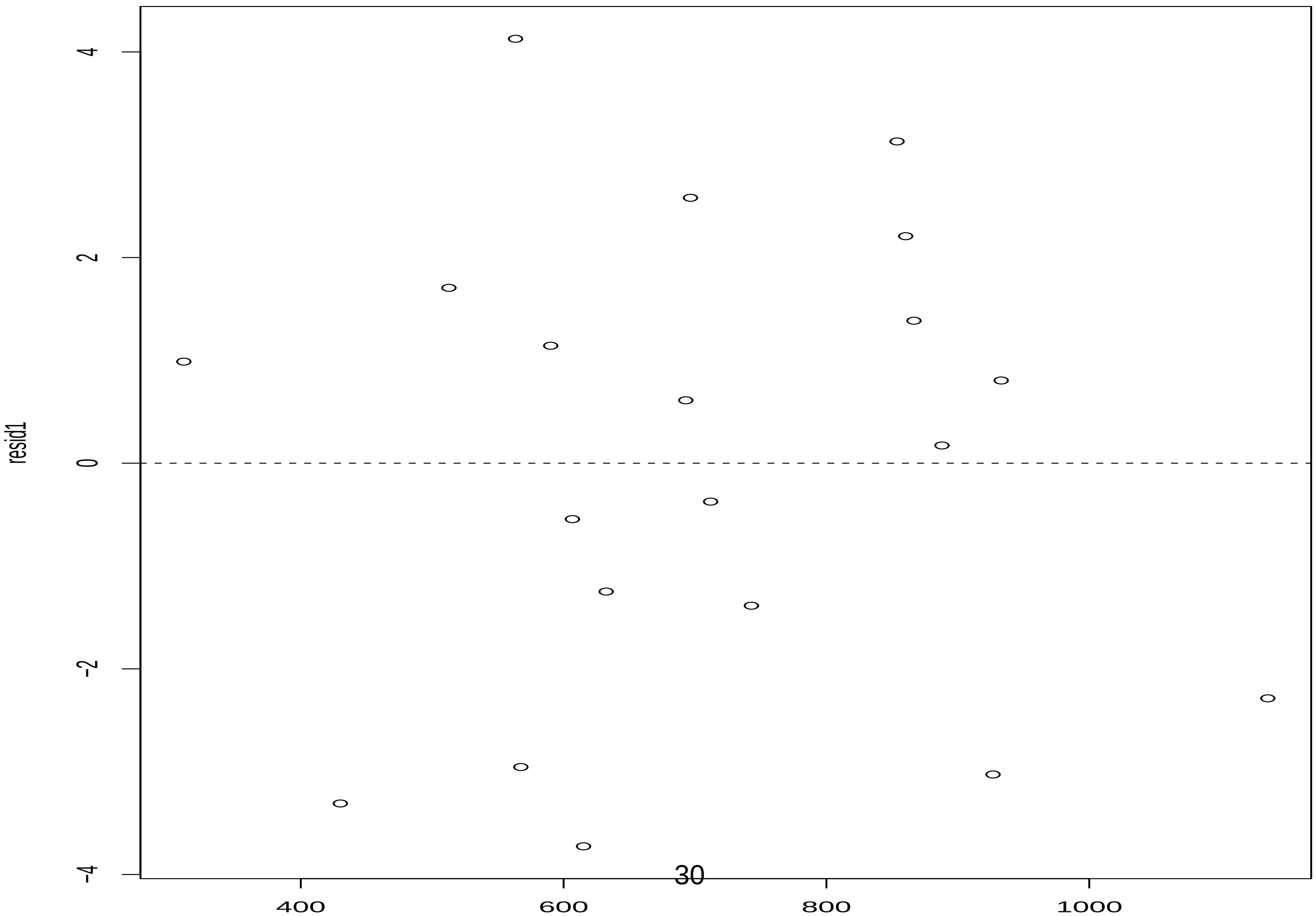
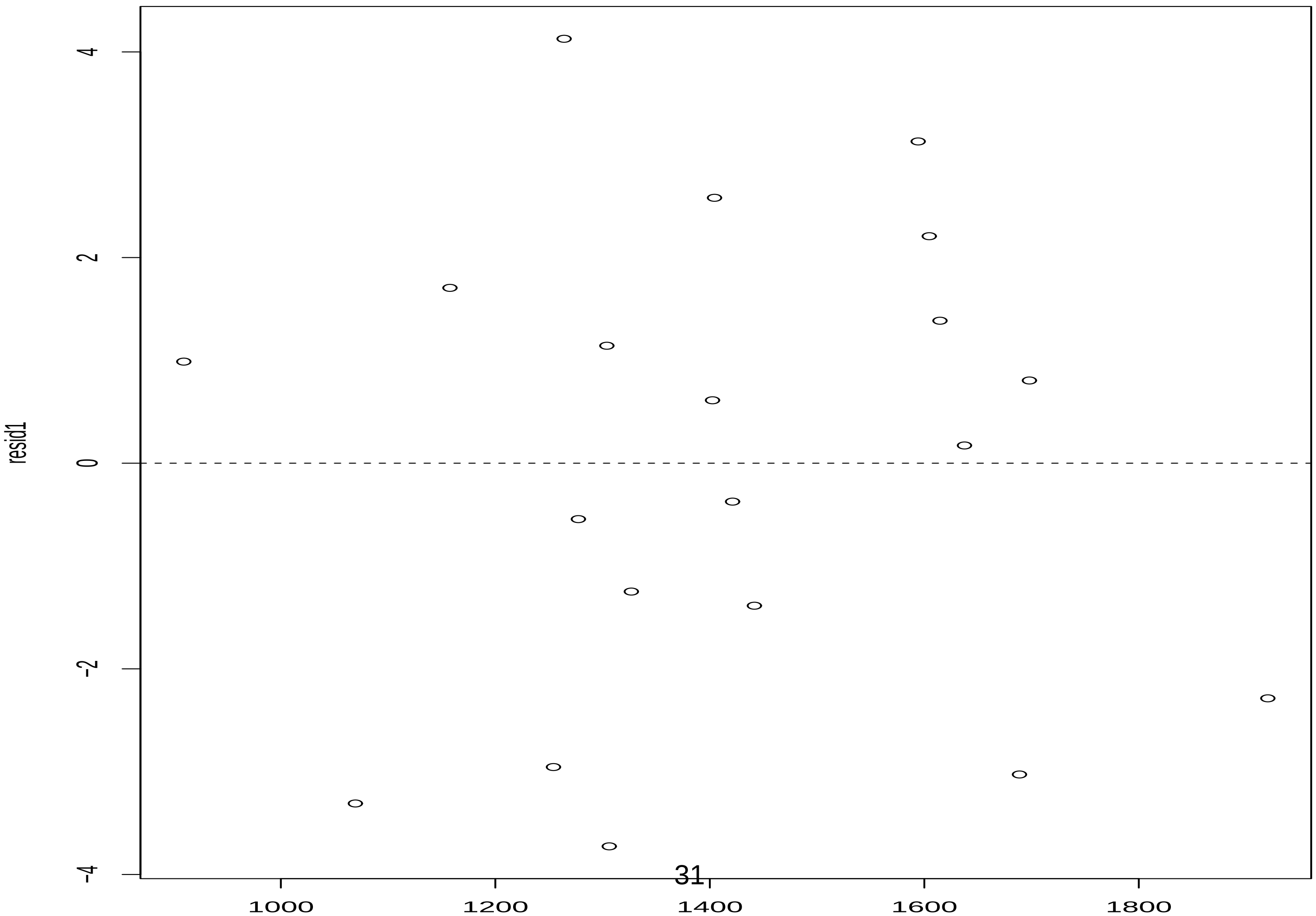


Figure 11: Bodyfat: residual from the additive model vs. interaction x2x3, didn't see a pattern, don't include



Comments:

- High multicollinearity may exist between some of the predictor variables and some of the interaction terms, as well as among some of the interaction terms. A partial remedy to improve computational accuracy is to center the predictor variables
- When the number of predictor variables is large, potential number of interaction terms become very large
- It is desirable to identify in advance, whenever possible, those interactions that are most likely to influence the response variable in important ways. In addition to utilizing a priori knowledge, one can plot the residuals for the additive regression model against the different interaction terms to determine which ones appear to be influential in affecting the response variable
- When the number of predictor variables is large, these plots may need to be limited to interaction terms involving those predictor variables that appear to be the most important on the basis of the initial fit of the additive regression model

Centering:

- Adding interaction terms to a regression model, high multicollinearities may exist between some of the predictor variables and some of the interaction terms, as well as among some of the interaction terms

- Remove or reduce the correlation between explanatory variables and their interactions by centering: subtracting the mean from each variable $x_{ik}^* = x_{ik} - \bar{x}_k$. Sometimes we rescale by standardizing (subtract the mean and divide by the standard deviation).

- ANOVA table after centering

Variable	Extra Sum of Squares
x_1	$SSR(x_1) = 352.27$
x_2	$SSR(x_2 \mid x_1) = 33.169$
x_3	$SSR(x_3 \mid x_1, x_2) = 11.546$
$x_1 x_2$	$SSR(x_1 x_2 \mid x_1, x_2, x_3) = 1.496$
$x_1 x_3$	$SSR(x_1 x_3 \mid x_1, x_2, x_3, x_1 x_2) = 2.704$
$x_2 x_3$	$SSR(x_2 x_3 \mid x_1, x_2, x_3, x_1 x_2, x_1 x_3) = 6.515$

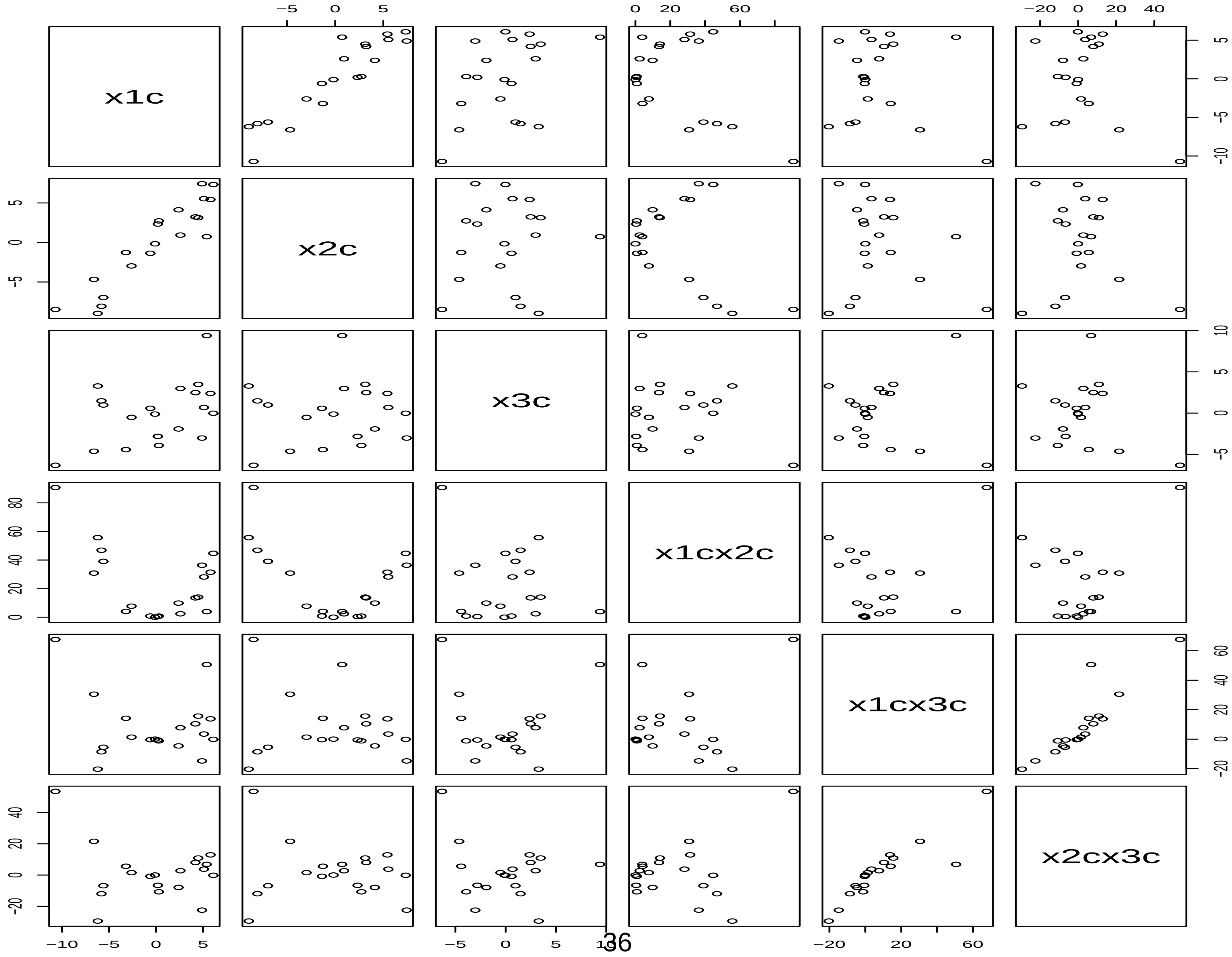
- Test whether any interaction terms are needed

$H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ v.s H_α: not all βs in H_0 equal zero.

$$\begin{aligned} F^* &= \frac{SSR(x_1x_2, x_1x_3, x_2x_3|x_1, x_2, x_3)/3}{MSE} \\ &= \frac{10.715/3}{6.745} \\ &= .53 \end{aligned}$$

For level of significance $\alpha = .05$, we require $F(.95; 3, 13) = 3.41$. Since $F^* = .53 < 3.41$, we conclude H_0, that the interaction terms are not needed in the regression model. The p-value of this test is .67.

Figure 12: Bodyfat: scatterplot for predictor variables after centering



8.3 Qualitative Predictors

- Many predictor variables of interest in business, economics, and the social and biological sciences are qualitative.

—-Gender: male, female

—-Purchase status: purchase, no purchase

—-Satisfaction with customer service:

(5-point scale - Very satisfied, Satisfied, Neutral/Not sure, Dissatisfied, Very dissatisfied)

Why include a qualitative independent variable?

- We are interested in the effect of a qualitative independent variable (for example: do men earn more than women?)
- We want to better predict/describe the dependent variable. We can make the errors smaller by including variables like gender, race, etc.

Example: An economist wished to relate the speed with which a particular insurance innovation is adopted (Y) to the size of the insurance firm (X_1) and the type of firm.

- Y: measured by the number of months elapsed between the time the first firm adopted the innovation and the time the given firm adopted the innovation
- X_1: size of firm, is quantitative, measured by the amount of total assets of the firm
- Type of firm, is qualitative and is composed of two classes

stock companies and mutual companies.

Note: in order that such a qualitative variable can be used in a regression model, indicator variables that take on the values 0 and 1 for the classes of the qualitative variable must be employed

Use two indicator variables:

$$X_2 = \begin{cases} 1 & \text{if stock company} \\ 0 & \text{otherwise} \end{cases} \qquad X_3 = \begin{cases} 1 & \text{if mutual company} \\ 0 & \text{otherwise} \end{cases}$$

A first-order model then would be the following:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

However, the columns of $\mathbf{X}$ are linearly dependent, rank of design matrix $\mathbf{X}$ is 3, which leads to computational difficulties.

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & 1 & 0 \\ 1 & X_{21} & 1 & 0 \\ 1 & X_{31} & 0 & 1 \\ 1 & X_{41} & 0 & 1 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & \sum_{i=1}^{4} X_{i1} & 2 & 2 \\ \sum_{i=1}^{4} X_{i1} & \sum_{i=1}^{4} X_{i1}^2 & \sum_{i=1}^{2} X_{i1} & \sum_{i=3}^{4} X_{i1} \\ 2 & \sum_{i=1}^{2} X_{i1} & 2 & 0 \\ 2 & \sum_{i=3}^{4} X_{i1} & 0 & 2 \end{bmatrix}$$

- The first column of the $\mathbf{X}'\mathbf{X}$ matrix equals the sum of the last two columns, so that the columns are linearly dependent

- $\mathbf{X}'\mathbf{X}$ doesn't have an inverse, no unique estimators of the regression coefficients can be found

- A simple way out of this difficulty is to drop one of the indicator variables

Suppose that we drop the indicator variable X_3, so regression model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \tag{15}$$

where: $X_{i1} =$ size of firm i and

$$X_{i2} = \begin{cases} 1 & \text{if company i is a stock company} \\ 0 & \text{otherwise} \end{cases}$$

Table 7: Data and Indicator Coding—-Insurance Innovation Example

Firm i	Number of months Elapsed Y_i	Size of Firm (million dollars) X_{i1}	Type of Firm	Indicator Code X_{i2}	$X_{i1}X_{i2}$
1	17	151	Mutual	0	0
2	26	92	Mutual	0	0
3	21	175	Mutual	0	0
4	30	31	Mutual	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
11	28	164	Stock	1	164
12	15	272	Stock	1	272
13	11	295	Stock	1	295
14	38	68	Stock	1	68
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

The response function for this regression model is

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{16}$$

$X_2 = 0$, $E(Y) = \beta_0 + \beta_1 X_1$ Mutual firms
$X_2 = 1$, $E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1$ Stock firms

- Mean time elapsed before the innovation is adopted, $E(Y)$, is a linear function of size of firm (X_1), with the same slope β_1 for both types of firms

- β_2 indicates how much higher (lower) the response function for stock firms is than the one for mutual firms, for any given size of firm

- In general, β_2 shows how much higher (lower) the mean response line is for the class coded 1 than the line for the class coded 0, for any given level of X_1

Table 8: Regression results for fit of regression model — insurance innovation example

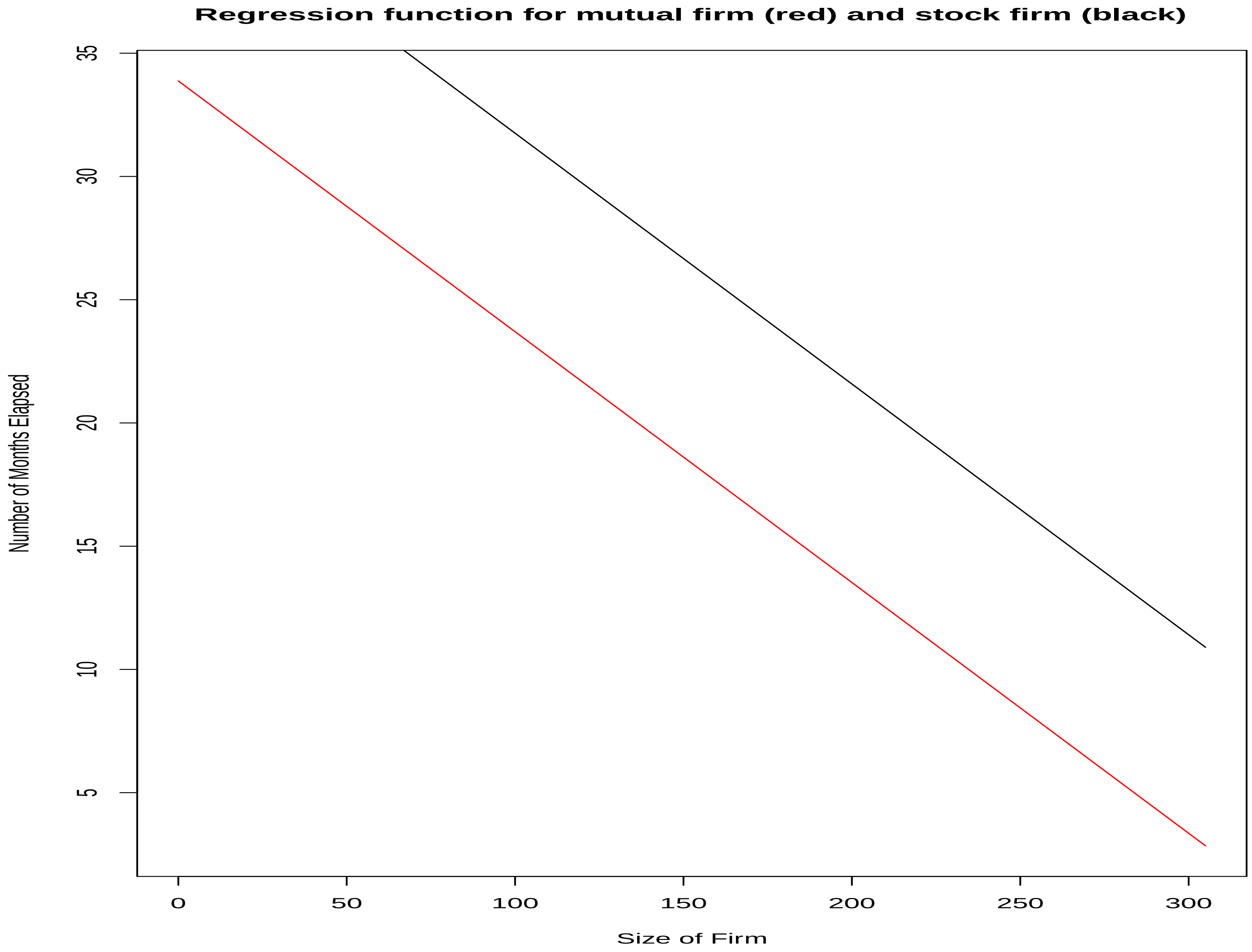
regression coefficient	estimates	se	t	p-value
(Intercept)	33.874069	1.813858	18.675	9.15e-13 ***
β_1	-0.101742	0.008891	-11.443	2.07e-09 ***
β_2	8.055469	1.459106	5.521	3.74e-05 ***

$$\hat{Y} = 33.87407 - .101742X_1 + 8.05547X_2$$

$$\hat{Y} = 33.87407 - .101742X_1$$ Mutual firms response function

$$\hat{Y} = (33.87407 + 8.05547) - .101742X_1$$ Stock firms response function

Figure 13: Regression lines: Mutual Firm and Stock Firm



regression coefficient	estimates	se	t	p-value
β_2	8.055469	1.459106	5.521	3.74e-05 ***

Interested in: the effect of type of firm (X_2) on the elapsed time for the innovation to be adopted.

- A formal test of $H_0 : \beta_2 = 0$ v.s $H_\alpha : \beta_2 \neq 0$ with significance 0.05 would lead to H_α, that type of firm has an effect.
- $n = 20$, $t(0.975, 17) = 2.110$

  CI of β_2 is: $8.05547 \pm 2.110 * 1.459106 = (4.98, 11.13)$

  —With 95% confidence, we conclude that stock companies tend to adopt the innovation somewhere between 5 and 11 months later, on the average, than mutual companies, for any given size of firm.

Comments:

- A qualitative variable with c classes will be represented by $c-1$ indicator variables, each taking on the values 0 and 1
- Indicator variables are frequently also called dummy variables or binary variables
- Models containing some quantitative and some qualitative explanatory variables, where the chief explanatory variables are quantitative and the qualitative variables are introduced primarily to reduce the variance of the error terms, are called *regression analysis*
- Models in which all explanatory variables are qualitative are called *analysis of variance models*
- Models containing some quantitative and some qualitative explanatory variables, where the chief explanatory variables are qualitative and the quantitative variables are introduced primarily to reduce the variance of the error terms, are called *analysis of covariance models*

Model Containing Interaction Effects

Introduce interaction term to the insurance innovation example

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

where:

$X_{i1} =$ size of firm i

$$X_{i2} = \begin{cases} 1 & \text{if company i is a stock company} \\ 0 & \text{otherwise} \end{cases}$$

The response function for this regression model is:

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

Meaning of Regression Coefficients:

For a mutual firm, $X_2 = 0$ and hence $X_1X_2 = 0$,

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3(0) = \beta_0 + \beta_1 X_1 \quad \text{Mutual firms}$$

For a stock firm, $X_2 = 1$ and hence $X_1X_2 = X_1$,

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1 = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 \quad \text{Stock firms}$$

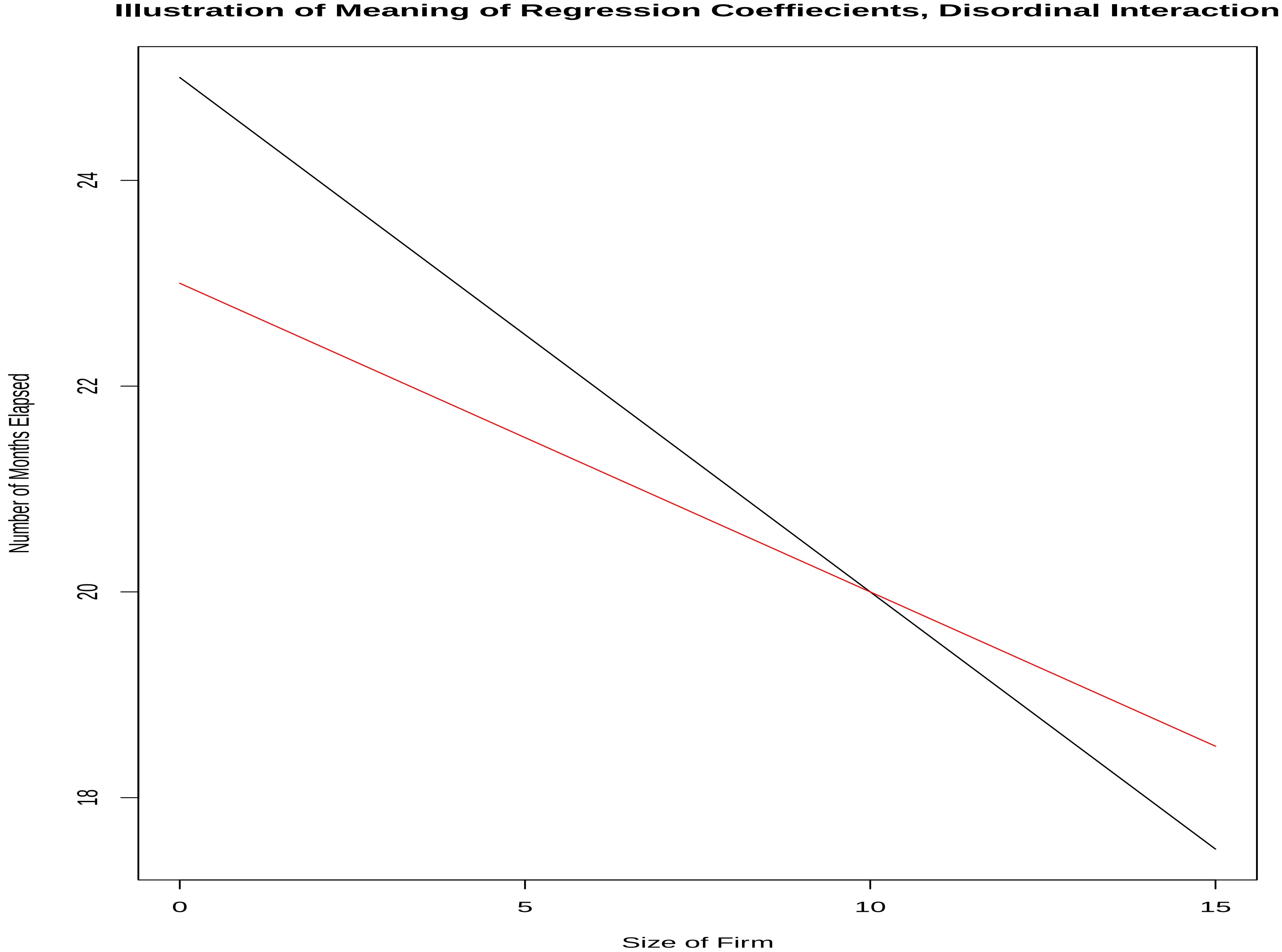
- β_2 indicates how much greater (smaller) is the Y intercept of the response function for the class coded 1 than that for the class coded 0.

- β_3 indicates how much greater (smaller) is the slope of the response function for the class coded 1 than that for the class coded 0.

- Effects of type of firm with interaction regression model depends on X_1, the size of the firm.

- Possible interaction: Disordinal interaction and ordinal interaction

Figure 14: Disordinal Interactions: Mutual Firm and Stock Firm



- For smaller firms, mutual firms tend to innovate more quickly, while for larger firms, stock firms tend to innovate more quickly

- When interaction effects are present, the effect of the qualitative predictor variable can only be studied by comparing the entire regression functions within the scope of the model for the different classes of the qualitative variable.

- When the response functions intersect within the scope of the model, the interaction is then said to be a disordinal interaction.

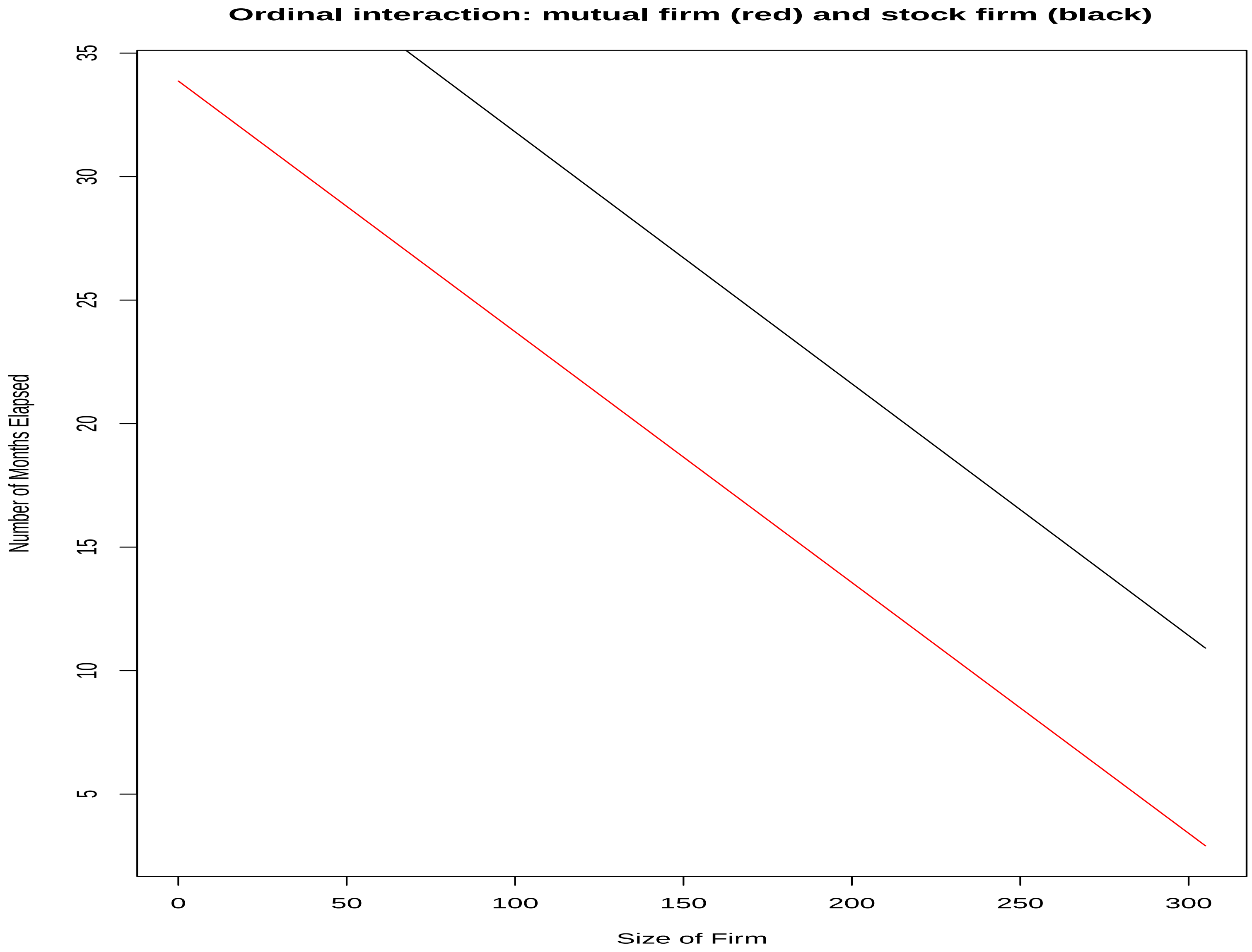
Table 9: Regression results for fit of interaction regression model — insurance innovation example

regression coefficient	estimates	se	t	p-value
(Intercept)	33.874069	2.44065	13.86	2.47e-13 ***
β_1	-0.10153	0.01305	-7.78	7.97e-07 ***
β_2	8.13125	3.65405	2.23	0.0408 *
β_3	-0.0004171	0.01833	-0.02	0.9821

$$\hat{y} = 33.874069 - 0.10153X_1 \quad \text{Mutual firms}$$

$$\begin{aligned} \hat{y} &= (33.874069 + 8.13125) + (-0.10153 - 0.0004171)X_1 \\ &= 42.005319 - 0.1019471X_1 \quad \text{Stock firms} \end{aligned}$$

Figure 15: Ordinal Interactions: Mutual Firm and Stock Firm



- Mutual firms tend to innovate more quickly than stock firms for all sizes of firms in the scope of the model
- When the response functions do not intersect within the scope of the model, the interaction is then said to be a ordinal interaction.

Recall:

Table 10: Regression results for fit of interaction regression model — insurance innovation example

regression coefficient	estimates	se	t	p-value
(Intercept)	33.874069	2.44065	13.86	2.47e-13 ***
β_1	-0.10153	0.01305	-7.78	7.97e-07 ***
β_2	8.13125	3.65405	2.23	0.0408 *
β_3	-0.0004171	0.01833	-0.02	0.9821

- whether the two regression functions have same slope

$$H_0 : \beta_3 = 0 \quad \text{v.s } H_\alpha : \beta_3 \neq 0$$

p-value = 0.9821, conclude that the two regression functions have the same slope, or there is no interaction effect, therefore, additive model should be adopted

- a test of whether the two regression functions are identical

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{v.s } H_\alpha \text{ not both } \beta_2 = 0 \text{ and } \beta_3 = 0$$

$$F^* = \frac{(SSE(R) - SSE(F))/2}{MSE(F)} = \frac{(492.63 - 176.38)/2}{11.02} = 14.344$$

Compared to $F(0.95, 2, 16)$ or p-value = 0.00027, reject H_0, conclude that the reduced model is not adequate, or the two regression functions are not identical.

Qualitative predictor with more than two classes

Consider regression of tool wear (y) on tool speed ($x1$) and tool model, where the latter is a qualitative variable with four classes ($M1, M2, M3, M4$). Require three indicator variables. Define

$$x_2 = \begin{cases} 1 & \text{if M1} \\ 0 & \text{otherwise} \end{cases} \quad x_3 = \begin{cases} 1 & \text{if M2} \\ 0 & \text{otherwise} \end{cases} \quad x_4 = \begin{cases} 1 & \text{if M3} \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$$

Tool model	$x1$	$x2$	$x3$	$x4$
$M1$	x_{i1}	1	0	0
$M2$	x_{i1}	0	1	0
$M3$	x_{i1}	0	0	1
$M4$	x_{i1}	0	0	0

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

$$E(y) = \beta_0 + \beta_1 x_1 \quad \text{Tool models} M_4$$

$$E(y) = (\beta_0 + \beta_2) + \beta_1 x_1 \quad \text{Tool models} M_1$$

$$E(y) = (\beta_0 + \beta_3) + \beta_1 x_1 \quad \text{Tool models} M_2$$

$$E(y) = (\beta_0 + \beta_4) + \beta_1 x_1 \quad \text{Tool models} M_3$$

- The regression of tool wear on tool speed is linear, with the same slope for all four tool models

- The coefficient β_2, β_3 and β_4 indicate, respectively, how much higher (lower) the response functions for tool models $M1, M2,$ and $M3$ are than the one for tool model $M4$, for any given level of tool speed