Chapter 8
Regression Models for Quantitative(量的) and Qualitative(質的) Predictors

許湘伶

Applied Linear Regression Models
(Kutner, Nachtsheim, Neter, Li)
Uses Polynomial Models

Polynomial regression models (多項式迴歸模型) for quantitative(量的) predictor variables

- the most frequently used curvilinear response models
- handled easily
- special case of the general LRM
Tow basic types

- The true curvilinear response function is a **polynomial function**
- The true curvilinear response function is **unknown**; a polynomial function is a **good approximation to the true function**

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.
Polynomial regression models: contain one, two, or more than two predictor variables

Each predictor variable: in various powers (次方)

**a second-order model with one predictor variable**

\[ Y_i = \beta_0 + \beta_1 (X_i - \bar{X}) + \beta_2 (X_i - \bar{X})^2 + \varepsilon_i \]

- the predictor variable is centered \((x_i = X_i - \bar{X})\):
  \[ \therefore X, X^2 \text{ often will be highly correlated} \]
- Centering: reduce the multicollinearity
- the quadratic response function: parabola (拋物線)

\[ E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 \]
8.1 Polynomial Regression Models

Figure: Examples of Second-Order Polynomial Response Functions.

(a) $E(Y) = 52 + 8x - 2x^2$

(b) $E(Y) = 18 - 8x + 2x^2$
8.1 Polynomial Regression Models

\[ E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 \]

Parameters:
- \( \beta_0 \): the mean response of \( Y \) when \( x = 0 \) (i.e. \( X = \bar{X} \))
- \( \beta_1 \): the linear effect coefficient
- \( \beta_{11} \): the quadratic effect coefficient
Third-order model with one predictor variable

\[ Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i \]

where \( x_i = X_i - \bar{X} \)

\[ E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 + \beta_{111} x^3 \]
Third-order model with one predictor variable (cont.)

Figure: Examples of Third-Order Polynomial Response Functions.
8.1 Polynomial Regression Models

**Higher orders with one predictor variable**

- Employed with special caution (謹慎)
- The interpretation of the coefficients becomes difficult
- The models may be highly erratic (無規律的; 不穩定的;) for interpolations (插補) and small extrapolations (外推 (法))
- A polynomial model of sufficiently high order can always be found to fit data containing no repeat observations perfectly.
  - Illustration: the fitted polynomial regression function for one predictor variable of order \( n - 1 \) will pass through all \( n \) observed \( Y \) values.
the second order with two predictor variables

The regression model:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \]

where: \( x_{i1} = X_{i1} - \bar{X}_1 \), \( x_{i2} = X_{i2} - \bar{X}_2 \)

The response function: (a conic section 圓錐形的截面)

\[ E\{ Y \} = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2}_{\text{linear comp.}} + \underbrace{\beta_{11} x_1^2 + \beta_{22} x_2^2}_{\text{quadratic comp.}} + \underbrace{\beta_{12} x_1 x_2}_{\text{cross-product}} \]

- \( \beta_{12} \): the interaction effect coefficient
- the second-order model with three predictor variables is similar
the second order with two predictor variables (cont.)

Figure: Examples $EY = 1,740 - 4x_1^2 - 3x_2^2 - 3x_1x_2$. (By mathematica)

maximum: $x_1 = 0; x_2 = 0$
the second order with two predictor variables (cont.)

```r
## Figure of 8.3
require(grDevices)
x1<-seq(-10,10,0.1)
x2<-seq(-10,10,0.1)
y <- function(x1, x2) {1740-4*x1^2-3*x2^2-3*x1*x2}
z <- outer(x1, x2, y)
persp(x1, x2, z, theta = -30, phi = 20, expand = 1,
col = "red",ticktype = "detailed",xlab="x1",ylab = "x2",
zlab = "E(y)"
)  
image(x1, x2, z, col  =terrain.colors(100))
contour(x1, x2, z, col = "blue", add = TRUE, method = "edge",
vfont = c("sans serif", "plain"))
```
Implementation of Polynomial Regression Models

- **Special cases** of the general linear regression model
- **Fitting** of polynomial models: no new problems
- All earlier results on fitting apply on making inferences.
Hierarchical Approach (階層式方式) to Fitting

Ideas to find an approximation to the true regression function:

1. often fit a second-order or third-order model

   \[ Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i \]

2. explore whether a lower-order model is adequate
   - Test \( \beta_{111} = 0 \)
   - Or test not both \( \beta_{11} = 0; \beta_{111} = 0 \)

3. The decomposition of \( SSR \) into extra sums of squares:

   \[ SSR(x); \quad SSR(x^2|x); \quad SSR(x^3|x, x^2) \]

4. Test whether \( \beta_{111} = 0 \): \( SSR(x^3|x, x^2) \)
   Test whether \( \beta_{11} = \beta_{111} = 0 \):
   \[ SSR(x^2, x^3|x) = SSR(x^2|x) + SSR(x^3|x, x^2) \]
Hierarchical Approach (階層式方式) to Fitting (cont.)

If a polynomial term of a given order is retained (保留), then all related terms of lower order are also retained in the model.

- providing more basic information about the shape of the response function

Wish to express the model in terms of the original variables:

\[
\hat{Y} = b_0 + b_1 x + b_{11} x^2 = b_0 + b_1 (X - \bar{X}) + b_{11} (X - \bar{X})^2
\]

\[
= (b_0 - b_1 \bar{X} + b_{11} \bar{X}^2) + (b_1 - 2b_{11} \bar{X}) X + b_{11} X^2
\]

\[
= b'_0 + b'_1 X + b'_{11} X^2
\]

- The fitted values and residuals for the regression function in terms of \(x\) or \(X\) are the same.

- Centered observations: to reduce potential calculation difficulties; multicollinearity;
Case Example: the life of power cell (電源電池; 蓄電池)

- Studied the effects of the charge rate and temperature on the life of a new type of power cell
- $X_1$: the charge rate at three levels (0.6, 1.0, 1.4 amperes 安培)
- $X_2$: the ambient (周遭的) temperature at three levels (10, 20, 30°C)
- $Y$: the life of the power cell
- don’t know the nature of the response function in the range of the factors studied
- fit the second-order polynomial regression model

\[
Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i
\]

\[
\Rightarrow E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2
\]
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

### Table: Data-Power Cells Example.

<table>
<thead>
<tr>
<th>Cell</th>
<th>Number of Cycles (thousands)</th>
<th>Charge Rate (thousands)</th>
<th>Temperature</th>
<th>Coded Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Yᵢ</td>
<td>Xᵢ₁</td>
<td>Xᵢ₂</td>
<td>xᵢ₁ xᵢ₂ xᵢ₁^2 xᵢ₂^2 xᵢ₁ xᵢ₂</td>
</tr>
<tr>
<td>1</td>
<td>150</td>
<td>0.6</td>
<td>10</td>
<td>-1 -1 1 1 1</td>
</tr>
<tr>
<td>2</td>
<td>86</td>
<td>1.0</td>
<td>10</td>
<td>0 -1 0 1 0</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>1.4</td>
<td>10</td>
<td>1 -1 1 1 -1</td>
</tr>
<tr>
<td>4</td>
<td>288</td>
<td>0.6</td>
<td>20</td>
<td>-1 0 1 0 0</td>
</tr>
<tr>
<td>5</td>
<td>157</td>
<td>1.0</td>
<td>20</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>6</td>
<td>131</td>
<td>1.0</td>
<td>20</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>7</td>
<td>184</td>
<td>1.0</td>
<td>20</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td>109</td>
<td>1.4</td>
<td>20</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>9</td>
<td>279</td>
<td>0.6</td>
<td>30</td>
<td>-1 1 1 1 -1</td>
</tr>
<tr>
<td>10</td>
<td>235</td>
<td>1.0</td>
<td>30</td>
<td>0 1 0 1 0</td>
</tr>
<tr>
<td>11</td>
<td>224</td>
<td>1.4</td>
<td>30</td>
<td>1 1 1 1 1</td>
</tr>
</tbody>
</table>

\[ \bar{X}_1 = 1.0\]
\[ \bar{X}_2 = 20\]

\[ x_{i1} = \frac{X_{i1} - \bar{X}_1}{0.4}; \quad x_{i2} = \frac{X_{i2} - \bar{X}_2}{10} \]
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

```r
ex<-read.table("CH08TA01.txt",header=F)
colnames(ex)<-c("Y","X1","X2")
attach(ex)
X1sqr<-X1^2
X2sqr<-X2^2
X1X2<-X1*X2
x1<-(ex$X1-mean(ex$X1))/0.4
x2<-(ex$X2-mean(ex$X2))/10
x1sqr<-x1^2
x2sqr<-x2^2
x1x2<-x1*x2
```
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

<table>
<thead>
<tr>
<th>Correlation between</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ and $X_1^2$: .991</td>
<td></td>
</tr>
<tr>
<td>$x_1$ and $x_1^2$: 0.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation between</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$ and $X_2^2$: .986</td>
<td></td>
</tr>
<tr>
<td>$x_2$ and $x_2^2$: 0.0</td>
<td></td>
</tr>
</tbody>
</table>

Interested in whether interaction effects and curvature effects are required in the model.

Fitted model:

$$\hat{Y} = 162.84 - 55.83x_1 + 75.50x_2 + 27.39x_1^2 - 10.61x_2^2 + 11.50x_1x_2$$
Case Example: the life of power cell (電池)

```r
> summary(lm(Y~x1+x2+x1sqr+x2sqr+x1x2))
Call:
  lm(formula = Y ~ x1 + x2 + x1sqr + x2sqr + x1x2)
Residuals:
   10 11 7.263 13.202
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept) 162.84      16.61  9.805   0.000188 ***
x1       -55.83       13.22 -4.224   0.008292 **
x2        75.50       13.22  5.712   0.002297 **
x1sqr     27.39       20.34  1.347   0.235856
x2sqr    -10.61       20.34 -0.521   0.624352
x1x2      11.50       16.19  0.710   0.509184
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 32.37 on 5 degrees of freedom
Multiple R-squared: 0.9135,    Adjusted R-squared: 0.8271
F-statistic: 10.57 on 5 and 5 DF,  p-value: 0.01086
```
8.1 Polynomial Regression Models

Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

Model: MODEL1
Dependent Variable: Y

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>66365.56140</td>
<td>11073.11228</td>
<td>10.566</td>
<td>0.0109</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>5240.43860</td>
<td>1048.08772</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>10</td>
<td>66606.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 32.37418 R-square 0.5135
Dep Mean 172.00000 Adj R-sq 0.5271
C.V. 18.82220

### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for Ho: Parameter=0 | Prob > |T|
|----------|----|--------------------|----------------|-----------------------|---------|
| INTERCEPT| 1  | 162.842105         | 16.60760542    | 9.805                 | 0.0002  |
| X1       | 1  | -55.833333         | 13.21670433    | -4.224                | 0.0083  |
| X2       | 1  | 75.500000          | 13.21670433    | 5.712                 | 0.0023  |
| X1SQ     | 1  | 27.394737          | 20.34007956    | 1.347                 | 0.2359  |
| X2SQ     | 1  | -10.605263         | 20.34007956    | -0.521                | 0.6244  |
| X1X2     | 1  | 11.500000          | 16.18709146    | 0.710                 | 0.5092  |

Variable DF Type I SS
| INTERCEPT | 1  | 325424 |
| X1        | 1  | 18704  |
| X2        | 1  | 34262  |
| X1SQ      | 1  | 1645.966667 |
| X2SQ      | 1  | 284.920870 |
| X1X2      | 1  | 529.000000 |
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

None of the plots suggest any gross inadequacies of the model.

The coefficient of correlation between the ordered residuals and their expected values: 0.974

Figure: Diagnostic Residual Plots-Power Cells Example.
8.1 Polynomial Regression Models

Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

- **Test of fit**: replications at $x_1 = 0, x_2 = 0$

  \[
  SSPE = 1,404.67 \quad (df = n - c = 2) \\
  SSLF = 3,835.77 \quad (df = c - p = 3) \\
  F^* = 1.82 \leq F(0.95; 3, 2) = 19.2
  \]

  ⇒ conclude that the second-order polynomial regression function is a good fit

- **Coefficient of Multiple Determination**: $R^2 = 0.9135$
  
  $R_a^2 = 0.8271$ (Figure 8.4)

  The variation in the lives of the power cells is reduced by about 91% when the first-order and second-order relations to the change rate and ambient temperature are utilized.
## Multiple regression: lack of fit

### Method 1

```r
fit <- lm(Y ~ x1 + x2 + x1sqr + x2sqr + x1x2)
exfactor = factor(c(seq(-4,-1), rep(0,3), seq(1,4)))
# fit full model
anova(fit, lm(Y ~ exfactor))
```

### Method 2

```r
library(alr3) # for lack of fit
pureErrorAnova(fit)
```
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1</td>
<td>1</td>
<td>18704</td>
<td>18704</td>
<td>26.6315</td>
<td>0.0356 *</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>34202</td>
<td>34202</td>
<td>48.6970</td>
<td>0.01992 *</td>
</tr>
<tr>
<td>x1sqr</td>
<td>1</td>
<td>1646</td>
<td>1646</td>
<td>2.3436</td>
<td>0.26546</td>
</tr>
<tr>
<td>x2sqr</td>
<td>1</td>
<td>285</td>
<td>285</td>
<td>0.4057</td>
<td>0.58935</td>
</tr>
<tr>
<td>x1x2</td>
<td>1</td>
<td>529</td>
<td>529</td>
<td>0.7532</td>
<td>0.47696</td>
</tr>
<tr>
<td>Residuals</td>
<td>5</td>
<td>5240</td>
<td>1048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of fit</td>
<td>3</td>
<td>3836</td>
<td>1279</td>
<td>1.8205</td>
<td>0.37378</td>
</tr>
<tr>
<td>Pure Error</td>
<td>2</td>
<td>1405</td>
<td>702</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

- **Partial $F$ Test**: whether a first-order model would be sufficient?

  \[ H_0 : \beta_{11} = \beta_{22} = \beta_{12} = 0 \]

  \[ H_a : \text{not all } \beta \text{s in } H_0 \text{ equal zero} \]

  \[ F^* = \frac{SSR(x_1^2, x_2^2, x_1x_2|x_1, x_2)}{3} \div MSE = \frac{2,459.9}{3} \div 1048.1 \]

  \[ = 0.78 \leq F(0.95; 3, 5) = 5.41 \]

  \[ \Rightarrow \text{conclude } H_0: \text{no curvature and interaction effects are needed} \]

### R Code

```r
> anova(lm(Y~x1+x2), fit)
Analysis of Variance Table
Model 1: Y ~ x1 + x2
Model 2: Y ~ x1 + x2 + x1sqr + x2sqr + x1x2
  Res.Df RSS Df Sum of Sq F Pr(>F)
1     8 7700.3
2     5 5240.4 3    2459.9 0.7823 0.5527
```
Case Example: the life of power cell (電池; 蓄電池) (cont.)

- First-order Model:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \]

\[ \Rightarrow \hat{Y} = 172.00 - 55.83 x_1 + 75.50 x_2 \]

\[ \Rightarrow \hat{Y} = 160.58 - 139.58 X_1 + 7.55 X_2 \text{ the original variables} \]
Case Example: the life of power cell (電池; 蓄電池) (cont.)

Figure: S-Plus Plot of Fitted Response Plane (8.19)-Power Cells Example.
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

- **Estimation of Regression Coefficients:**
  - to estimate the linear effect of the two predictor variables with a 90% family confidence coefficients by the Bonferroni method. \((g = 2)\)

\[
B = t \left(1 - 0.1/2(2)\right) = 2.306
\]

\[
s\{b_1'\} = \left(\frac{1}{0.4}\right)s\{b_1\} = 31.68
\]

\[
s\{b_2'\} = \left(\frac{1}{10}\right)s\{b_2\} = 1.267
\]

\[
\Rightarrow -212.6 \leq \beta_1 \leq -66.5 \quad 4.6 \leq \beta_2 \leq 10.5
\]
Case Example: the life of power cell (電源電池; 蓄電池) (cont.)

\[ -212.6 \leq \beta_1 \leq -66.5 \quad 4.6 \leq \beta_2 \leq 10.5 \]

With confidence 0.90, we conclude that the mean number of charge/discharge cycles before failure decreases by 66 to 213 cycles with a unit increase in the charge rate for given ambient temperature, and increase by 5 to 10 cycles with a unit increase of ambient temperature for given charge rate.

\[
\begin{align*}
\text{(Intercept)} & \quad 64.617930 & \quad 256.54874 \\
X1 & \quad -212.602056 & \quad -66.56461 \\
X2 & \quad 4.629251 & \quad 10.47075
\end{align*}
\]
Some further Comments on polynomial regression

Drawbacks (缺點): polynomial models

- Such models can be more expensive in degrees of freedom than alternative nonlinear models or linear models with transformed variables.

- An alternative to using centered variables is to use orthogonal polynomials. Orthogonal polynomials are uncorrelated. (Some computer packages used orthogonal polynomials)

- Sometimes a quadratic response function is fitted for the purpose of establishing the linearity of the response function when repeat observations are not available.
Interaction Regression Models

Interaction effects:

- A regression model with \( p - 1 \) predictor variables contains additive effects:
  \[
  E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2
  \]
  \[
  f_1(X_1) + f_2(X_2) + \cdots + f_{p-1}(X_{p-1})
  \]

- \( f_i, i = 1, \ldots, p - 1 \): any functions, not necessarily simple one

Illustration:

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

⇒ the effects of \( X_1, X_2 \) on \( Y \) are additive
a cross-product term: modelling the interaction effect of two predictor variables on the response variable; ex: $\beta_3 X_1 X_2$

The cross-product term: called an interaction term; a linear-by-linear or a bilinear interaction term
Interaction Regression Models (cont.)

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}X_{i2} + \varepsilon_i \]

- The meaning of $\beta_1, \beta_2$ is not the same as that given earlier ($\because \beta_3 X_{i1}X_{i2}$)
- $\beta_1, \beta_2$ no longer indicate the change in the mean response with a unit increase of the predictor variable with the other predictor variable held constant.
- the change in the mean response with a unit increase in $X_1$ when $X_2$ is held constant:

\[ \frac{\partial E\{Y\}}{\partial X_1} = \beta_1 + \beta_3 X_2 \] (depend on $X_2$)
Interaction Regression Models (cont.)

- (a): parallel; condition effect plot
- (b): the slope \((2.5, 3.5)\) of the response function vs. \(X_1\) differ for \(X_2 = 1, 3\)

\[
\begin{align*}
E\{Y\} &= 10 + 2X_1 + 5X_2 \\
E\{Y\} &= 10 + 2X_1 + 5X_2 + .5X_1X_2
\end{align*}
\]
Interaction Regression Models (cont.)

- $\beta_1, \beta_2$ are positive: the interaction effect is called a reinforcement or synergistic type ($\beta_3$: positive)
- an interference or antagonistic type

\[
E\{Y\} = 10 + 2X_1 + 5X_2
\]

\[
E\{Y\} = 10 + 2X_1 + 5X_2 - 0.5X_1X_2
\]
shape of response function

\[ E\{Y\} = 10 + 2X_1 + 5X_2 \]
8.2 Interaction Regression Models

shape of response function (cont.)

\[ E\{Y\} = 10 + 2X_1 + 5X_2 + .5X_1X_2 \]
shape of response function (cont.)

\[ E\{Y\} = 10 + 2X_1 + 5X_2 - 0.5X_1X_2 \]
8.2 Interaction Regression Models

\[ E\{Y\} = 65 + 3X_1 + 4X_2 - 10X_1^2 - 15X_2^2 + 35X_1X_2 \]

**Figure**: Response Surfaces and Contour Curves for Curvilinear Regression Model with Interaction Effect-Quick Bread Volume Example.
Curvilinear (曲線的) Effects

\[ E\{Y\} = 65 + 3X_1 + 4X_2 - 10X_1^2 - 15X_2^2 + 35X_1X_2 \]

Figure: Conditional Effect Plot for Curvilinear Regression Model with Interaction Effect-Quick Bread Volume Example.
Implementation of interaction regression models

Considerations:

- high multicollinearities: center the predictor variables
  \[ x_{ik} = X_{ik} - \bar{X}_k \]

- The number of predictor variables is large \(\Rightarrow\) the potential number of interaction terms becomes very large

  utilizing *a priori* knowledge

plot the residuals for the additive regression model vs. the different interaction terms to determine which ones appear to be influential
Body fat example

- Model: Three predictor variables

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1}X_{i2} + \beta_5 X_{i1}X_{i3} + \beta_6 X_{i2}X_{i3} + \varepsilon_i \]

- Some of the predictor variables are highly correlated with some of the interaction terms

- Centered variables:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1}x_{i2} + \beta_5 x_{i1}x_{i3} + \beta_6 x_{i2}x_{i3} + \varepsilon_i \]

\[ \Rightarrow \hat{Y} = 20.53 + 3.438x_1 - 2.095x_2 - 1.616x_3 + 0.00888x_1x_2 - 0.0847 \]

\[ MSE = 6.745 \]
8.2 Interaction Regression Models

Implementation of interaction regression models (cont.)

\[
H_0 : \beta_4 = \beta_5 = \beta_6 = 0
\]

\[
H_a : \text{not all } \beta\text{s in } H_0 \text{ equal zero}
\]

\[
\Rightarrow F^* = \frac{SSR(x_1x_2, x_1x_3, x_2x_3|x_1, x_2, x_3)}{3} \div \text{MSE}
\]

\[
= 0.53 \leq F(0.95; 3, 13) = 3.41
\]

\[
\Rightarrow \text{conclude } H_0: \text{ the interaction terms are not needed in the regression model. (} P\text{-value} = 0.67)\]
Qualitative Predictors

Field: (Qualitative variables) business, economics, the social, biological sciences

Ex:
- gender (M,F);
- purchase status: purchase, no purchase
- disability status: not; partial; fully

A study of innovation in the insurance industry
- related the speed with which a particular insurance innovation is adopted ($Y$)
- the size of the insurance firm ($X_1$)
- the type of the firm
Qualitative Predictors (質的) Predictors (cont.)

Qualitative predictor with two classes:

- indicator variable: take on 0 and 1
- Ex: two indicator variables $X_2, X_3$

$$X_2 = \begin{cases} 
1 & \text{if stock company (股份公司)} \\
0 & \text{otherwise}
\end{cases}$$

$$X_3 = \begin{cases} 
1 & \text{if mutual company (互助公司)} \\
0 & \text{otherwise}
\end{cases}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$
Qualitative Predictors (cont.)

- Intuitive approach of setting up an indicator variable: leads to computation difficulties
- Design matrix $\mathbf{X}$ with $n = 4$:

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

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

- $\mathbf{X}$: the first column = the sum of the last two columns $\Rightarrow$ linearly dependent
- $\mathbf{X}'\mathbf{X}$: no inverse
Simple way out of the difficulty: **drop one indicator variable**

**Principle**

A qualitative variable with *c* classes will be represented by *c − 1* indicator variables, each taking on the values 0 and 1.

**Indicator variables:** called *dummy variables* or *binary variables*.

(drop \(X_3\))

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i
\]

\[
E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2
\]
Meaning of the regression coefficients:

\[ E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \]

- \( X_2 = 0 \):
  \[ E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(0) = \beta_0 + \beta_1 X_1 \] Mutual firms

- \( X_2 = 1 \):
  \[ E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(1) = (\beta_0 + \beta_2) + \beta_1 X_1 \] Stock firms
Figure: Illustration of Meaning of Regression Coefficient for Regression Model (8.33) with Indicator Variable $X_2$-Insurance Innovation Example.
## Table: Data and Indicator Coding-Insurance Innovation Example.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$Y_i$</th>
<th>$X_{i1}$</th>
<th>$X_{i2}$</th>
<th>$X_{i1}X_{i2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>151</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>92</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>175</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>31</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>104</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>277</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>210</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>120</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>290</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>238</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>28</td>
<td>164</td>
<td>1</td>
<td>164</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>272</td>
<td>1</td>
<td>272</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>295</td>
<td>1</td>
<td>295</td>
</tr>
<tr>
<td>14</td>
<td>38</td>
<td>68</td>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
<td>85</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>224</td>
<td>1</td>
<td>224</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>166</td>
<td>1</td>
<td>166</td>
</tr>
<tr>
<td>18</td>
<td>13</td>
<td>305</td>
<td>1</td>
<td>305</td>
</tr>
<tr>
<td>19</td>
<td>30</td>
<td>124</td>
<td>1</td>
<td>124</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
<td>246</td>
<td>1</td>
<td>246</td>
</tr>
</tbody>
</table>
Insurance innovation example (cont.)

- studied 10 mutual firms and 10 stock firms
- The fitted of regression model:

\[ \hat{Y} = 33.87407 - 0.10174X_1 + 8.05547X_2 \]

- Interested in the effect of type of firm (\(X_2\))
- a 95% confidence interval for \(\beta_2\):

\[ 4.98 \leq \beta_2 \leq 11.13 \quad (t(0.975; 17) = 2.110) \]

- Test:

\[ H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_a : \beta_2 \neq 0 \]

\[ \Rightarrow \text{lead to } H_a: \text{type of firm has an effect (} \alpha = 0.05) \]
Insurance innovation example (cont.)

Figure: Regression Results for Fit of Regression Model (8.33)-Insurance Innovation Example.

<table>
<thead>
<tr>
<th>Regression Coefficient</th>
<th>Estimated Regression Coefficient</th>
<th>Estimated Standard Deviation</th>
<th>t*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>33.87407</td>
<td>1.81386</td>
<td>18.68</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1.10174</td>
<td>0.00889</td>
<td>-11.44</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>8.05547</td>
<td>1.45911</td>
<td>5.52</td>
</tr>
</tbody>
</table>

(b) Analysis of Variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1,504.41</td>
<td>2</td>
<td>752.20</td>
</tr>
<tr>
<td>Error</td>
<td>176.39</td>
<td>17</td>
<td>10.38</td>
</tr>
<tr>
<td>Total</td>
<td>1,680.80</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>
Insurance innovation example (cont.)

Figure: Fitted Regression Function for Regression Model (8.33) Indicator Innovation Example.
Insurance innovation example (cont.)

```r
ex<-read.table("CH08TA02.txt",header=F)
colnames(ex)<-c("Y","X1","X2")
attach(ex)
cX2<-asfactor(X2)
fit<-lm(Y~X1 +cX2)
coef <- coefficients(fit)
plot( Y ~ X1, pch=(15+X2),col = as.character(2+X2))
abline(coef["(Intercept)"] + coef["cX21"], coef["X1"], col = 3)
abline(coef["(Intercept)"] , coef["X1"], col = 2)
legend(35,10,c("Stock firm","Mutul firm"),col=c(3,2), lwd=1,
 lty=c(1,1), pch=c(16,15))
```
Insurance innovation example (cont.)

Why did we not simply fit separate regressions for stock firms and mutual firms?

1. The model assumes equal slopes and the same constant error term variance for each type of firm.

2. The common slope $\beta_1$ can best be estimated by pooling the two types of firms.

3. Other inferences, $\beta_0, \beta_2$, can be made more precisely by working with one regression model containing an indicator variable ($\therefore$ more df associated with $MSE$).
More than two classes

- The regression of tool wear (磨損) \((Y)\) on tool speed \(X_1\) and tool model (型號) (qualitative: M1, M2, M3, M4)

- Require three indicator variables:

\[
X_2 = \begin{cases} 
1 & \text{if tool model M1} \\
0 & \text{otherwise}
\end{cases} \quad X_3 = \begin{cases} 
1 & \text{if tool model M2} \\
0 & \text{otherwise}
\end{cases} \\
X_4 = \begin{cases} 
1 & \text{if tool model M3} \\
0 & \text{otherwise}
\end{cases}
\]
More than two classes (cont.)

First-order model

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \]

\[ \Rightarrow E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \]

\[ \Leftrightarrow \text{Tool model M4: } E\{Y\} = \beta_0 + \beta_1 X_1 \]

Tool model M1: \( E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \)

Tool model M2: \( E\{Y\} = (\beta_0 + \beta_3) + \beta_1 X_1 \)

Tool model M3: \( E\{Y\} = (\beta_0 + \beta_4) + \beta_1 X_1 \)

<table>
<thead>
<tr>
<th>Tool Model</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>( X_{i1} )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M2</td>
<td>( X_{i1} )</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M3</td>
<td>( X_{i1} )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M4</td>
<td>( X_{i1} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
the regression of tool wear on tool speed is **linear**, with the **same slope** for all four tool models

$\beta_2, \beta_3, \beta_4$: how much higher (lower) the response functions for toll models (M1,M2,M3) are than the one for M4

always compared with the class for which $X_2 = X_3 = X_4 = 0$
Figure: Illustration of Regression Model (8.36) - Tool Wear Example.

More than two classes (cont.)
More than two classes (cont.)

- wish to **estimate differential effects** other than against tool models M4:
  
  \[ \Rightarrow \text{estimate differences between regression coefficients} \]
  
  (ex: \( \beta_4 - \beta_3 \))

- The point estimator is \( b_4 - b_3 \)

- the estimated variance of this estimator:

\[
s^2\{ b_4 - b_3 \} = s^2\{ b_4 \} + s^2\{ b_3 \} - 2s\{ b_4, b_3 \}
\]
8.3 Qualitative Predictors

**Time Series Applications**

- Economists; business analysis
- Using indicator variables:
  - year: peacetime; wartime (戰時)

\[ Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \epsilon_t, \quad t = 1, \ldots, n \]

where \( X_{t1} = \text{income} \)

\[ X_{t2} = \begin{cases} 
1 & \text{if period } t \text{ peace time} \\
0 & \text{otherwise} 
\end{cases} \]

- monthly or quarterly: seasonal effect
- Time series data: susceptible (容許…的) to correlated error terms (Chap. 12)
Some Considerations in Using Indicator Variables

Allocated codes: arbitrary; other numbers;
- define a metric for the classes of the qualitative variable

<table>
<thead>
<tr>
<th>Class</th>
<th>$X_1$</th>
<th>$E{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent user</td>
<td>3</td>
<td>$E{Y} = \beta_0 + \beta_1(3) = \beta_0 + 3\beta_1$</td>
</tr>
<tr>
<td>Occasional user</td>
<td>2</td>
<td>$E{Y} = \beta_0 + \beta_1(2) = \beta_0 + 2\beta_1$</td>
</tr>
<tr>
<td>Nonuser</td>
<td>1</td>
<td>$E{Y} = \beta_0 + \beta_1(1) = \beta_0 + \beta_1$</td>
</tr>
</tbody>
</table>

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

- The key implication:
  $$E\{Y|\text{frequent user}\} - E\{Y|\text{occasional user}\}$$
  $$= E\{Y|\text{occasional user}\} - E\{Y|\text{nonuser}\} = \beta_1$$
Some Considerations in Using Indicator Variables (cont.)

Indicator variables: make no assumptions about the spacing of the classes;

<table>
<thead>
<tr>
<th>Class</th>
<th>$X_1$</th>
<th>$X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent user</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Occasional user</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Nonuser</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

$\beta_1$: $E\{Y|frequent\ user\} - E\{Y|nonuser\}$

$\beta_2$: $E\{Y|ocasional\ user\} - E\{Y|nonuser\}$

- $\beta_1 - \beta_2$: the differential effect between frequent user and occasional user
- if $\beta_1 = 2\beta_2 \Rightarrow$ equal spacing between the three classes
Indicator variables can be used even if the predictor variable is quantitative.

- transformed by grouping into classes
- age: 21, 21-34, 35-49

Information about the original variable may be thrown away.

Additional parameters into the model: reducing the df associated with $MSE$. 

Some Considerations in Using Indicator Variables (cont.)

Other codings:

- First coding:

\[
X_{i2} = \begin{cases} 
1 & \text{if stock company} \\
-1 & \text{if mutual company}
\end{cases}
\]

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i
\]

\[
E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{Stock firms}
\]

\[
E\{Y\} = (\beta_0 - \beta_2) + \beta_1 X_1 \quad \text{Mutual firms}
\]

⇒ \beta_0: "average" intercept of the regression line

- two regression lines are the same:

\[
H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_a : \beta_2 \neq 0
\]
Some Considerations in Using Indicator Variables (cont.)

- Second coding:
  - using **indicator variables** for each of the \( c \) classes of the qualitative variable
  - drop the intercept term
    \[ Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \]

  where: \( X_{i1} = \) size of firm; \( X_{i2} = \begin{cases} 1 & \text{if stock company} \\ 0 & \text{otherwise} \end{cases} \)

  \( X_{i3} = \begin{cases} 1 & \text{if mutual company} \\ 0 & \text{otherwise} \end{cases} \)

  \( \Rightarrow E\{Y\} = \beta_2 + \beta_1 X_1 \quad \text{Stock firms} \)

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

- The same regression line: \( H_0 : \beta_2 = \beta_3 \) vs. \( H_a : \beta_2 \neq \beta_3 \)
Modeling Interactions between Quantitative and Qualitative Predictors

- the possibility of interaction effects: the size of firm and type of firm

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}X_{i2} + \varepsilon_i \]

where \( X_{i1} = \text{size of firm} \)

\[ X_{i2} = \begin{cases} 
1 & \text{if stock company} \\
0 & \text{otherwise} 
\end{cases} \]

\[ \Rightarrow E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \]

\[ \Leftrightarrow E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (0) + \beta_3 (0) = \beta_0 + \beta_1 X_1 \text{ Mutual firms} \]

\[ E\{Y\} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 \text{ Stock firms} \]
Modeling Interactions between Quantitative and Qualitative Predictors (cont.)

Disordinal interaction (非次序性交互作用)

- Both the intercept and the slope differ for the two classed in the response function.
- The effect of the qualitative predictor variable can be studied only by comparing the regression functions within the scope of the model for the different classed of the qualitative variable.
Modeling Interactions between Quantitative and Qualitative Predictors (cont.)

ordinal interaction (次序性交互作用)

Figure: Another Illustration of Regression Model (8.49) with Indicator Variable $X_2$ and Interaction Term- Insurance Innovation Example.
Example: insurance innovation

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \]

**Figure**: Regression Result for Fit of Regression Model (8.49) with Interaction Term - Insurance Innovation Example.
Example: insurance innovation (cont.)

\[ H_0 : \beta_3 = 0 \text{ vs. } H_a : \beta_3 \neq 0 \Rightarrow |t^*| = 0.02 < t(0.975; 16) \]

\[ \Rightarrow \text{no interaction effects} \]

```r
> summary(lm(Y~X1+X2+X1*X2))
Call:
  lm(formula = Y ~ X1 + X2 + X1 * X2)
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 ***
     X2     8.1312501  3.6540517  2.225 0.0408   *
X1:X2    -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
```
More Complex Models

- two or more of the predictor are qualitative

\[ X_2 = \begin{cases} 
1 & \text{if firm incorporated} \\
0 & \text{otherwise} 
\end{cases} \]

\[ X_3 = \begin{cases} 
1 & \text{if quality of sales management high} \\
0 & \text{otherwise} 
\end{cases} \]

**first-order:** \[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \]

**interaction:** \[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1}X_{i2} + \beta_5 X_{i1}X_{i3} + \beta_6 X_{i2}X_{i3} + \epsilon_i \]

<table>
<thead>
<tr>
<th>Type of Firm</th>
<th>Quality of Sales Management</th>
<th>Response Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporated</td>
<td>High</td>
<td>[ E{Y} = (\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_4 + \beta_5)X_1 ]</td>
</tr>
<tr>
<td>Not incorporated</td>
<td>High</td>
<td>[ E{Y} = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)X_1 ]</td>
</tr>
<tr>
<td>Incorporated</td>
<td>Low</td>
<td>[ E{Y} = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)X_1 ]</td>
</tr>
<tr>
<td>Not incorporated</td>
<td>Low</td>
<td>[ E{Y} = \beta_0 + \beta_1X_1 ]</td>
</tr>
</tbody>
</table>
More Complex Models (cont.)

- Qualitative predictor variables only
  \[ Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \]

- All explanatory variables are qualitative: *analysis of variance (ANOVA變異數分析) models*

- Some qualitative & some qualitative explanatory variables: *analysis of covariance (ANCOVA共變異分析) models*
Comparison of Two or More Regression Functions

Three examples

- two production lines for making soap bars (肥皂)
- family income study from urban (城市的) and rural (農村的) areas
- instrument calibration (儀器口徑測定) study
production lines for making soap bar

- Two production lines ($X_2$): $X_1 =$ production line speed & $Y =$ the amount of the scrap (廢料)
- Linear but not the same for the two production lines: the same slopes & the different height
- A formal test is desired to determine whether or not the two regression lines are identical.
family income study

- modeled by linear regression
- wish to compare whether, at given income level, urban and rural families tend to save the same amount—i.e., whether the slopes of the two regression lines are the same
instrument calibration (儀器口徑測定) study

- Two instruments were constructed
- relation between gauge readings (計量儀器讀數) & actual pressures
- If the two regression lines are the same, a single calibration schedule can be developed for the two instruments; otherwise, two different calibration schedules will be required.