CHAPTER 8

MULTIPLE REGRESSION

8.4 For each of two models give the following points

(a) Centering predictors: **(5 points)**
(b) Model interpretation: **(10 points)**
(c) Model assessment: **(10 points)**
(d) Plots: **(10 points)**

Also, give **(10 points)** for model comparison. Thus the total points on this problem are **(80 points)**.

As an example of how the problem might be solved, two possible models (using centered predictors CPH and CDAYS) are:

- **Model 1 (ADDITIVE)** (Plots in Figures 1 and 2)
  o Fitted model:
  $$\widehat{LEAD} = 15.97 - 5.02 \cdot CPH + 0.56 \cdot CDAYS.$$
  o $R^2 = 0.5409$, $R_a^2 = 0.5235$, $MSE = 25.72$.
  o The intercept, 15.97, is the predicted lead level at mean PH and DAYS.
  o The change in predicted lead per unit change in CPH is -5.02, and the change in predicted lead per unit change in CDAYS is 0.56.
  o There is a slight curvature in the plot of residuals versus fitted values. Plots of studentized residuals look reasonable.

- **Model 2 (INTERACTION)** (Plots in Figures 3 and 4)
  o Fitted model:
  $$\widehat{LEAD} = 15.97 - 5.02 \cdot CPH + 0.56 \cdot CDAYS - 0.50 \cdot CPH \cdot CDAYS.$$
  o $R^2 = 0.6568$, $R_a^2 = 0.6370$, $MSE = 19.59$.
  o The intercept, 15.97, is the predicted lead level at mean PH and DAYS.
  o The change in predicted lead per unit change in CPH is

  $$\frac{\partial \, \widehat{LEAD}}{\partial \, CPH} = -5.02 - 0.50 \cdot CDAYS,$$

  and the change in predicted lead per unit change in CDAYS is

  $$\frac{\partial \, \widehat{LEAD}}{\partial \, CDAYS} = 0.56 - 0.50 \cdot CPH.$$

  o Plots of studentized residuals look reasonable.

  Model 2 is preferred because of R^2 R_a^2 and MSE.

8.6 $R^2 = 0.6568$.

- **(10 points)** 65.68% of the variation in LEAD is explained by the model: $R^2 = SSR/SSTO = 1949.8326/2968.7347 = 0.6568$.
- **(10 points)** The model reduces the uncertainty in predicting LEAD by 65.68%: $R^2 = 1 - SSE/SSTO = 1 - 1018.9021/2968.7347 = 0.6568$.

8.8 **(5 points)** The VIFs all equal 1.0, so there is no multicollinearity at all. The previous conclusions are unchanged.

8.12 **(5 points)** The overall model is significant, yet none of the regressors is. This seems a paradox. The explanation is multicollinearity, as shown by the large VIFs. I would try centering the predictors.

LEAD	=	CPH	CDAYS
Response Distribution:		Normal	
Link Function:		Identity	

Model Equation						
LEAD	=	15.9661	-	5.0180 CPH	+	0.5603 CDAYS

Summary of Fit			
Mean of Response	15.9661	R-Square	0.5409
Root MSE	5.0712	Adj R-Sq	0.5235

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	2	1605.7037	802.8519	31.2180	0.0001
Error	53	1363.0310	25.7176		
C Total	55	2968.7347			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
CPH	1	647.4925	647.4925	25.1771	0.0001
CDAYS	1	958.2112	958.2112	37.2590	0.0001

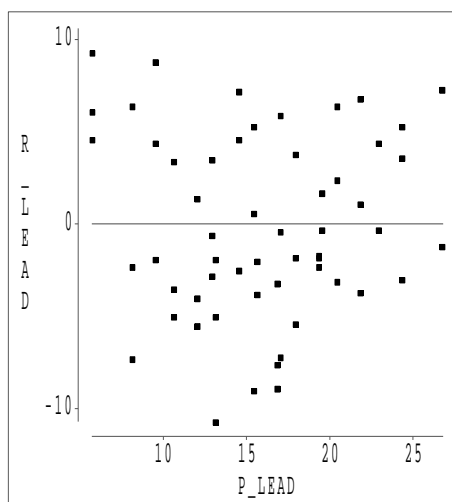| Parameter Estimates							
Variable	DF	Estimate	Std Error	T Stat	Prob > |T|	Tolerance	Var Inflation
INTERCEPT	1	15.9661	0.6777	23.5602	0.0001	.	0
CPH	1	-5.0180	1.0001	-5.0177	0.0001	1.0000	1.0000
CDAYS	1	0.5603	0.0918	6.1040	0.0001	1.0000	1.0000

Figure 1: *SAS/INSIGHT output, additive model, lead data.*
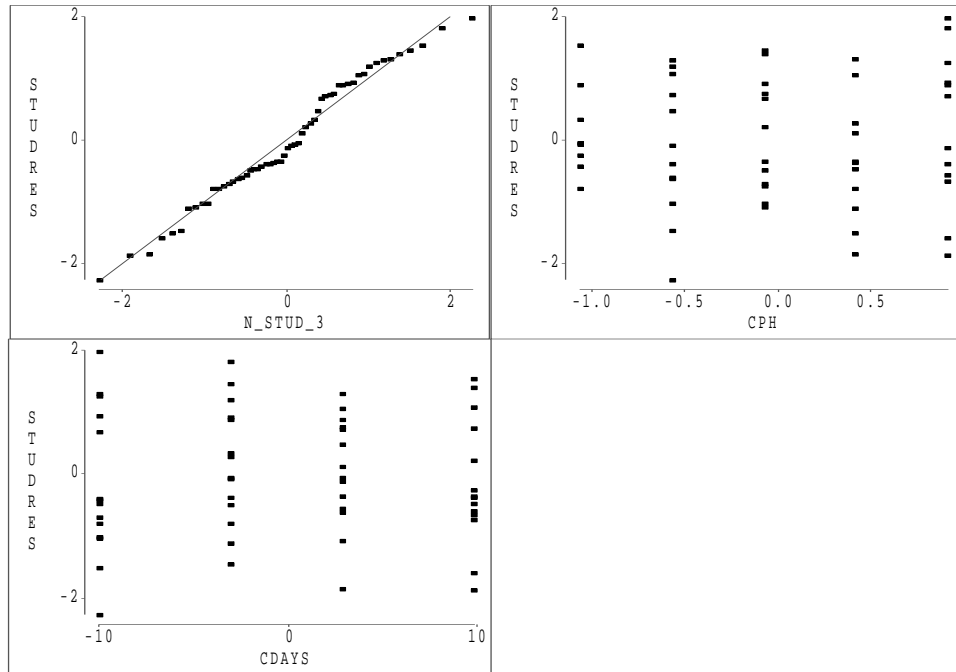
Figure 2: *Studentized residual plots, additive model, lead data.*

```
 ┌─────────────────────────────────────────┐
 │ LEAD    =   CPH  CDAYS  CPH*CDAYS         │
 ├─────────────────────────────────────────┤
 │ Response Distribution:      Normal        │
 │ Link Function:        Identity            │
 └─────────────────────────────────────────┘
```

Parameter Information	
Parameter	Variable
1	INTERCEPT
2	CPH
3	CDAYS
4	CPH*CDAYS

Model Equation
LEAD   =   15.9661   -   5.0180   CPH   +   0.5603   CDAYS   -   0.4955   P_4

Summary of Fit			
Mean of Response	15.9661	R-Square	0.6568
Root MSE	4.4265	Adj R-Sq	0.6370

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	3	1949.8326	649.9442	33.1701	0.0001
Error	52	1018.9021	19.5943		
C Total	55	2968.7347			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
CPH	1	647.4925	647.4925	33.0450	0.0001
CDAYS	1	958.2112	958.2112	48.9026	0.0001
CPH*CDAYS	1	344.1289	344.1289	17.5627	0.0001

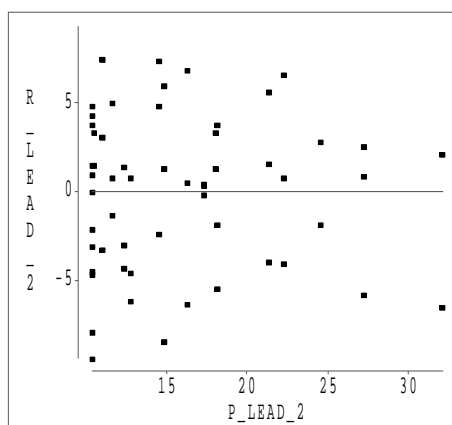| Parameter Estimates							
Variable	DF	Estimate	Std Error	T Stat	Prob > |T|	Tolerance	Var Inflation
INTERCEPT	1	15.9661	0.5915	26.9917	0.0001	.	0
CPH	1	-5.0180	0.8729	-5.7485	0.0001	1.0000	1.0000
CDAYS	1	0.5603	0.0801	6.9930	0.0001	1.0000	1.0000
CPH*CDAYS	1	-0.4955	0.1182	-4.1908	0.0001	1.0000	1.0000

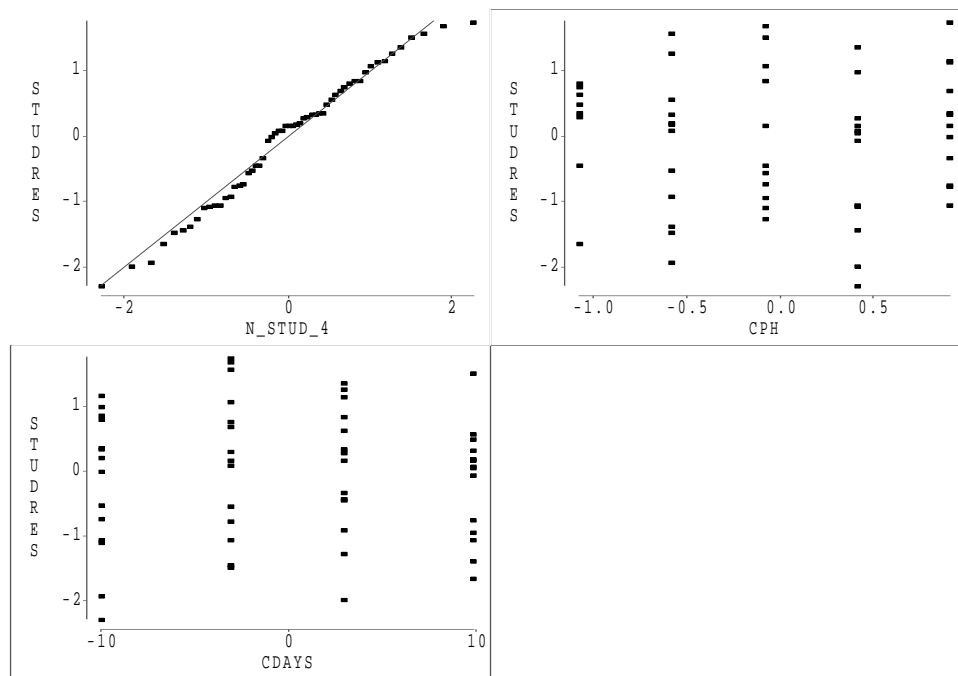Figure 3: *SAS/INSIGHT output, interaction model, lead data.*

Figure 4: *Studentized residual plots, interaction model, lead data.*