Assignment for Applied Regression Analysis

朱强强

17064001

Applied Statistics

3.1

a. From the table below we can conclude that
$y = -1.80837 + 0.00360x_2 + 0.19386x_7 - 0.00482x_8$.
But the coefficient of intercept is not significant.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.80837	7.90086	-0.23	0.8209
x2	1	0.00360	0.00069500	5.18	<.0001
x7	1	0.19396	0.08823	2.20	0.0378
x8	1	-0.00482	0.00128	-3.77	0.0009

b. We know that the p_value < 0.05, so the regression is reasonable.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	257.09428	85.69809	29.44	<.0001
Error	24	69.87000	2.91125		
Corrected Total	27	326.96429			

c. The last column of the first table shows the results of hypothesis testing for regression coefficients. We know that x_2 contributes the most to the regression model, followed by x_8 ,and finally x_7.

d. $R^2 = 0.7683, R^2_{Adj} = 0.7596$.

Root MSE	1.70624	R-Square	0.7863
Dependent Mean	6.96429	Adj R-Sq	0.7596
Coeff Var	24.49984		

e. From the table we can get the same conclusion as the t test for β_7 calculated in part c above.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x8		1	0.5447	0.5447	27.1368	31.10	<.0001
2	x2		2	0.1986	0.7433	6.8324	19.34	0.0002
3	x7		3	0.0430	0.7863	4.0000	4.83	0.0378

3.2

From the result computed by SAS, we know that square of the simple correlation coefficient between the observed values y_i and the fitted values $\hat{y}_i$ is 0.7924, which very nearly equals R^2 (0.7863) calculated before.

Pearson Correlation Coefficients, N = 28 Prob > |r| under H0: Rho=0		
	y	y_fit
y	1.00000	0.05208 0.7924
y_fit	0.05208 0.7924	1.00000

3.3

a. From the result calculated by SAS we can know that a 95% CI on β_7 is [0.01186, 0.37607].

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-1.80837	7.90086	-0.23	0.8209	-18.11494	14.49820
x2	1	0.00360	0.00069500	5.18	<.0001	0.00216	0.00503
x7	1	0.19396	0.08823	2.20	0.0378	0.01186	0.37607
x8	1	-0.00482	0.00128	-3.77	0.0009	-0.00745	-0.00218

b. A 95% CI on the mean number of games won by a team when $x_2 = 2300, x_7 = 56.0, x_8 = 2100$ is [6.4362, 7.9966].

Output Statistics						
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		Residual
1	.	7.2164	0.3780	6.4362	7.9966	.

3.4

a. From the table below we can know that the p_value calculated < 0.05, so the regression is significant.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	179.06619	89.53309	15.13	<.0001
Error	25	147.89810	5.91592		
Corrected Total	27	326.96429			

b. $R^2 = 0.5477, R^2_{Adj} = 0.5115$.

Root MSE	2.43227	R-Square	0.5477
Dependent Mean	6.96429	Adj R-Sq	0.5115
Coeff Var	34.92486		

These two figures are both smaller that those calculated in Problem 3.1, which indicates that the regressor in Problem 3.1 is better. So x_2 can make a significant contribution to a better regression model.

c. From the result calculated by SAS we can know that a 95% CI on β_7 is [-0.19716, 0.29391].

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	17.94432	9.86248	1.82	0.0808	-2.36785	38.25649
x7	1	0.04837	0.11922	0.41	0.6884	-0.19716	0.29391
x8	1	-0.00654	0.00176	-3.72	0.0010	-0.01016	-0.00292

A 95% CI on the mean number of games won by a team when $x_2 = 2300, x_7 = 56.0, x_8 = 2100$ is [5.8286, 8.0238].

Output Statistics						
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		Residual
1	.	6.9262	0.5329	5.8286	8.0238	.

3.29

we know that

$$\mathbf{X}=\begin{bmatrix}1 & x_{11} & x_{12} & \cdots & x_{1k}\\ 1 & x_{21} & x_{22} & \cdots & x_{2k}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk}\end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1}=\frac{1}{S_{xx}}\begin{pmatrix}\sum x_i^2/n & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}$$

so

$$\begin{aligned}h_{ii}&=(1\quad x_i)\left[\frac{1}{S_{xx}}\begin{pmatrix}\sum x_i^2/n & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}\right]\begin{pmatrix}1\\ x_i\end{pmatrix}\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2}{n}-x_i\bar{x}-x_i\bar{x}+x_i^2\right]\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2}{n}+(x_i-\bar{x})^2-\bar{x}^2\right]\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2-n\bar{x}^2}{n}+(x_i-\bar{x})^2\right]\\ &=\frac{1}{n}+\frac{(x_i-\bar{x})^2}{S_{xx}}\end{aligned}$$

and

$$\begin{aligned}h_{ij}&=(1\quad x_i)\left[\frac{1}{S_{xx}}\begin{pmatrix}\sum x_i^2/n & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}\right]\begin{pmatrix}1\\ x_j\end{pmatrix}\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2}{n}-x_i\bar{x}-x_j\bar{x}+x_ix_j\right]\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2}{n}+(x_i-\bar{x})(x_j-\bar{x})-\bar{x}^2\right]\\ &=\left[\frac{1}{S_{xx}}\right]\left[\frac{\sum x_i^2-n\bar{x}^2}{n}+(x_i-\bar{x})(x_j-\bar{x})\right]\\ &=\frac{1}{n}+\frac{(x_i-\bar{x})(x_j-\bar{x})}{S_{xx}}\end{aligned}$$

From the these expressions we can easily know that these quantities will increase as x_i moves farther from $\bar{x}$.

3.30

$$\begin{aligned}\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{y}&=(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\\ &=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\\ &=\boldsymbol{\beta}+\boldsymbol{R}\boldsymbol{\varepsilon}\\ \boldsymbol{R}&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\end{aligned}$$

3.31

$$\begin{aligned}\boldsymbol{e} &= (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y} \\ &= (\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{X\beta}+\boldsymbol{\varepsilon}) \\ &= [\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X\beta}+(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\varepsilon} \\ &= (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\varepsilon}\end{aligned}$$

3.39

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \cdots & 1 \end{bmatrix}$$

The j^{th} VIF is the j^{th} diagonal element of $(\mathbf{W}'\mathbf{W})^{-1}$.

$$(\mathbf{W}'\mathbf{W})^{-1} = \frac{1}{1-R_j^2}$$