Test Your Understanding 14

Baseball statistics guru Bill James claims that the expected winning percentage of any team is a function of the ratio of the total runs scored by the team to the total runs scored by its opponents, which we will call RATIO. In particular, he asserts that if P represents the winning percentage, then

$$E(P|\text{RATIO}) = -0.2222 + 0.9802\text{RATIO} - 0.2575\text{RATIO}^2.$$

a. Does the intercept have any meaning for this model? Explain.

   Solution: *The intercept would only have meaning if $RATIO = 0$ makes sense. It is unlikely that a team would score no runs in a season. In addition, even if $RATIO = 0$ makes sense, the value of the intercept is negative, which makes no sense as an expected winning percentage.*

b. Joe Bob says that 0.9802 is the change in expected winning percentage per unit change in RATIO when the other regressor is held constant. Comment.

   Solution: *Joe Bob is wrong. $RATIO^2$ cannot remain constant while RATIO changes.*

c. Interpret this model.

   Solution: *As we have seen, the intercept does not have a meaning. The estimated change in $E(P|RATIO)$ per unit change in RATIO is*

$$\frac{\partial}{\partial RATIO}E(P|RATIO) = 0.9802 - 0.515RATIO.$$

Test Your Understanding 15

Bill James got his model from a least squares fit performed on a set of data. SAS/INSIGHT output for the fit is shown in the two attached figures. Evaluate the fit of the model.

Solution: *The model explains over 85% of the variation in winning percentage ($R^2 = 0.85$). The plot of residuals versus fitted values shows no evidence of outliers or lack of fit. The histogram and normal quantile plot of the studentized residuals do not give reason to question the assumption of normality.*

Test Your Understanding 16

Two MLR models fit to the same set of data result in the following:

	Model 1	Model 2
Number of Regressors	2	3
R^2	0.9321	0.9525
R_a^2	0.8868	0.8813

Everything else being equal, which model would you prefer? Why?

Solution: *I would prefer model 1, since it is more parsimonious (2 regressors versus 3), and does not lose much explanatory power (it has a higher R_a^2 than model 2).*

Test Your Understanding 17

A MLR with 3 regressors is fit to 27 data values. The correlation between the response and the fitted values is 0.9714. Conduct the F test of $H_0: \beta_1 = \beta_2 = \ldots = \beta_q = 0$ at the 0.01 level of significance.

PCT = RATIO RATIO*RATIO
Response Distribution: Normal
Link Function: Identity

Parameter Information	
Parameter	Variable
1	INTERCEPT
2	RATIO
3	RATIO*RATIO

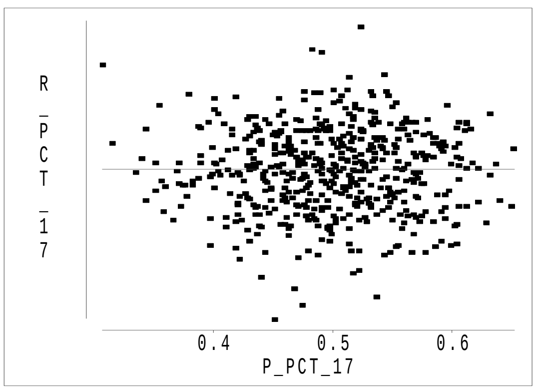
Model Equation
PCT = - 0.2222 + 0.9802 RATIO - 0.2575 P_3

Summary of Fit			
Mean of Response	0.5000	R-Square	0.8651
Root MSE	0.0249	Adj R-Sq	0.8646

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	2	2.1371	1.0685	1721.2980	0.0001
Error	537	0.3334	0.0006		
C Total	539	2.4704			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
RATIO	1	0.0810	0.0810	130.5102	0.0001
RATIO*RATIO	1	0.0237	0.0237	38.1078	0.0001

Parameter Estimates							
Variable	DF	Estimate	Std Error	T Stat	Prob >|T|	Tolerance	Var Inflation
INTERCEPT	1	-0.2222	0.0437	-5.0869	0.0001	.	0
RATIO	1	0.9802	0.0858	11.4241	0.0001	0.0082	122.2932
RATIO*RATIO	1	-0.2575	0.0417	-6.1732	0.0001	0.0082	122.2932



Solution:

$$F^* = \frac{n-q-1}{q}\frac{R^2}{1-R^2} = \frac{27-3-1}{3}\frac{0.9714^2}{1-0.9714^2} = 128.3.$$

$F_{3,23,0.99} = 4.765$. *Since* $F^* > F_{3,23,0.99}$, *we reject* H_0 *at the* 0.01 *significance level.*

Test Your Understanding 18

A MLR is fit to a set of data. The estimated standard error for the mean response a specified set of regressor values $x_1, x_2, \ldots, x_q$, is 12.96, while the estimated standard error of prediction of a future observation at the same set of regressor values is 19.46. What is the MSE from the model fit?

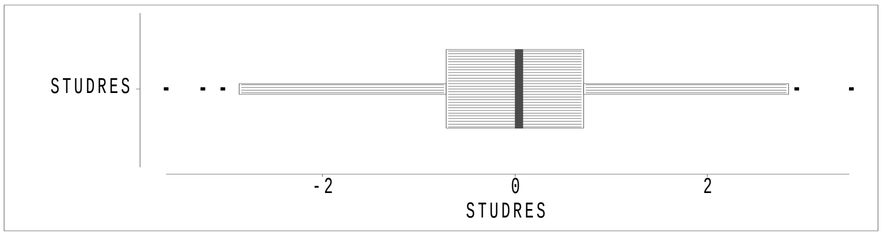
Solution:

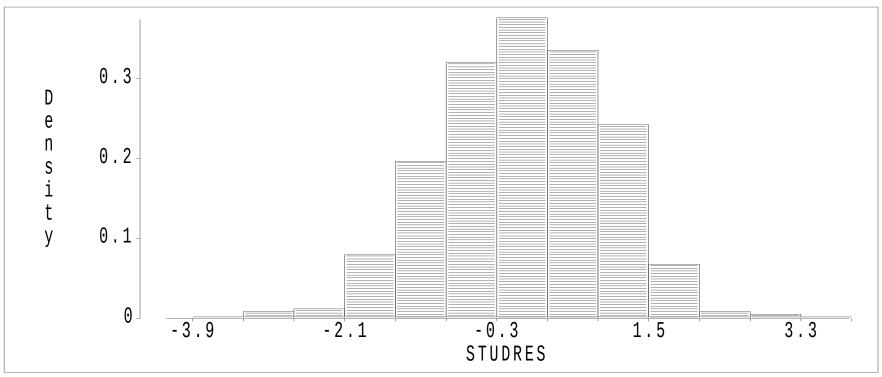
$$19.46 = \hat{\sigma}(Y_{new} - \hat{Y}_{new}) = \sqrt{MSE + 12.96^2},$$

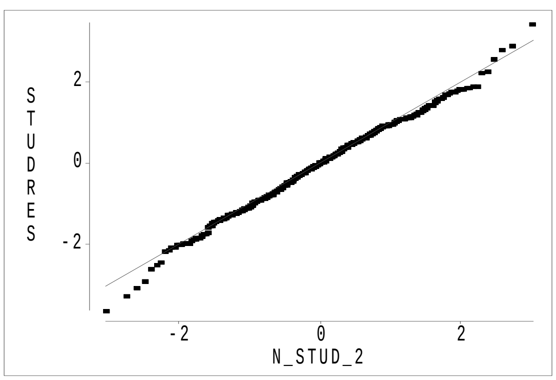
so

$$MSE = 19.46^2 - 12.96^2 = 210.73.$$

STUDRES







Moments

N	540.0000	Sum Wgts	540.0000
Mean	-3.262E-05	Sum	-0.0176
Std Dev	1.0029	Variance	1.0059
Skewness	-0.1562	Kurtosis	0.2935
USS	542.1644	CSS	542.1644
CV	-3074594.2	Std Mean	0.0432

Quantiles

100%	Max	3.4876	99.0%	2.2542
75%	Q3	0.7127	97.5%	1.8421
50%	Med	0.0416	95.0%	1.5606
25%	Q1	-0.7203	90.0%	1.1706
0%	Min	-3.6214	10.0%	-1.2374
	Range	7.1089	5.0%	-1.6996
	Q3-Q1	1.4331	2.5%	-1.9616
	Mode	-3.6214	1.0%	-2.4824

QQ Ref Line

Line	Intercept	Slope
	0	1.0000