

Example 5.2

A power company wants to use daily high temperature, x , to model daily peak power load, y , during the summer months when demand is greatest.

Although the company expects peak load to increase as the temperature increases, the *rate* of increase in $E(y)$ might not remain constant as x increases. For example, a 1-unit increase in high temperature from 100 °F to 101 °F might result in a larger increase in power demand than would a 1-unit increase from 80 °F to 81 °F.

Therefore, the company postulates that the model for $E(y)$ will include a second-order (quadratic) term and, possibly, a third-order (cubic) term.

A random sample of 25 summer days is selected and both the peak load (measured in megawatts) and high temperature (in degrees) recorded for each day. The data are listed in Table 5.1.

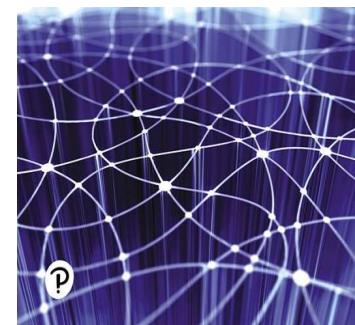
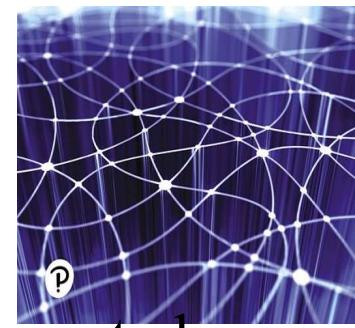


Table 5.1 Power load data

Temperature °F	Peak Load megawatts	Temperature °F	Peak Load megawatts	Temperature °F	Peak Load megawatts
94	136.0	106	178.2	76	100.9
96	131.7	67	101.6	68	96.3
95	140.7	71	92.5	92	135.1
108	189.3	100	151.9	100	143.6
67	96.5	79	106.2	85	111.4
88	116.4	97	153.2	89	116.5
89	118.5	98	150.1	74	103.9
84	113.4	87	114.7	86	105.1
90	132.0				



Example 5.2 (ctn)

(a) Construct a scatterplot for the data. What type of model is suggested by the plot?

[Exercise Q1]

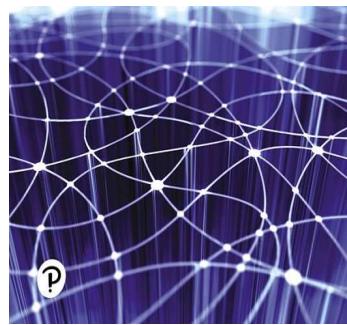
(b) Fit the third-order model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, to the data. Is there evidence that the cubic term, $\beta_3 x^3$, contributes information for the prediction of peak power load? Test at $\alpha = .05$.

[Exercise Q2]

(c) Fit the second-order model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$, to the data. Test the hypothesis that the power load increases at an increasing rate with temperature. Use $\alpha = .05$.

[Exercise Q3]

(d) Give the prediction equation for the second-order model, part c. Are you satisfied with using this model to predict peak power loads?

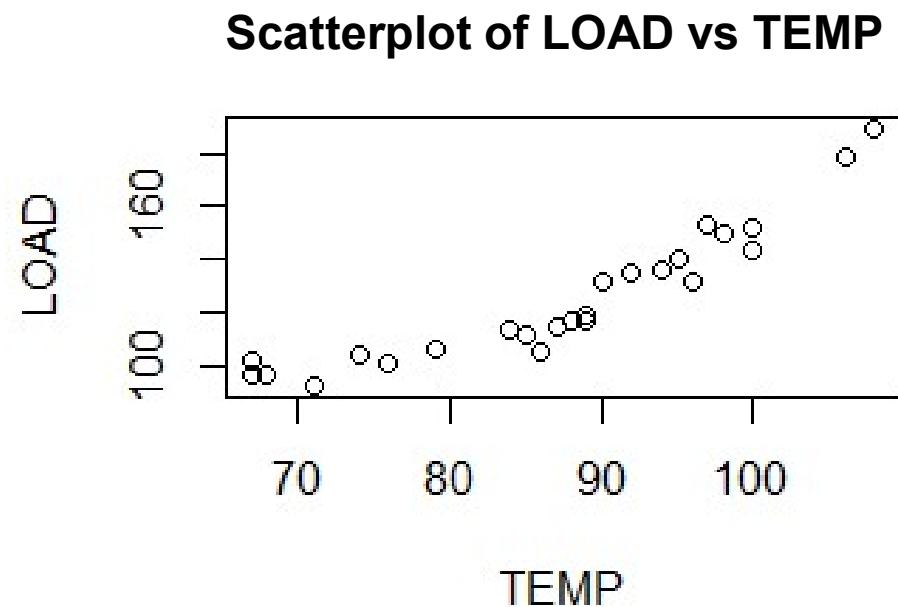


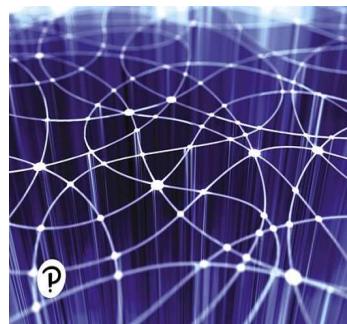
Example 5.2 (ctn)

Solution:

(a) The scatterplot shows a nonlinear, upward-curving trend, which indicates that a **second order model** would likely fit the data well.

Figure 5.5 R
scatterplot for power
load data





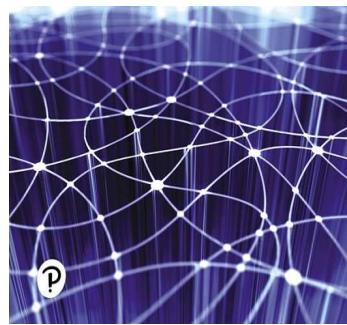
R output -- third_order_model

```
third_order_model <- lm(LOAD ~ TEMP + I(TEMP^2) + I(TEMP^3))
anova_alt(third_order_model)

## Analysis of Variance Table
##
##          Df    SS   MS      F      P
## Source  3 15012.2 5004.1 165.36 9.1368e-15
## Error   21  635.5   30.3
## Total   24 15647.7  652.0

summary(third_order_model)

##
## Call:
## lm(formula = LOAD ~ TEMP + I(TEMP^2) + I(TEMP^3))
##
## Residuals:
##       Min     1Q     Median      3Q     Max 
## -10.3229 -2.1941 -0.1422  3.3026  9.7775 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 331.25268035 477.11146572  0.694   0.495    
## TEMP        -6.39191245  16.79082602 -0.381   0.707    
## I(TEMP^2)     0.03775397  0.19451185  0.194   0.848    
## I(TEMP^3)     0.000008432  0.00074260  0.114   0.911    
## 
## Residual standard error: 5.501 on 21 degrees of freedom
## Multiple R-squared:  0.9594, Adjusted R-squared:  0.9536 
## F-statistic: 165.4 on 3 and 21 DF,  p-value: 9.137e-15
```



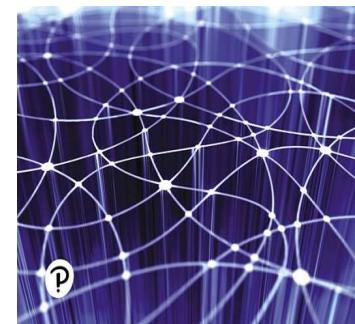
R output -- second_order_model

```
second_order_model <- lm(LOAD ~ TEMP + I(TEMP^2))
anova_alt(second_order_model)

## Analysis of Variance Table
##
##          Df    SS   MS      F    P
## Source    2 15011.8 7505.9 259.69 4.9908e-16
## Error   22  635.9   28.9
## Total    24 15647.7  652.0

summary(second_order_model)

##
## Call:
## lm(formula = LOAD ~ TEMP + I(TEMP^2))
##
## Residuals:
##       Min     1Q   Median     3Q    Max
## -10.4291 -2.1779 -0.0156  3.1759  9.6489
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 385.048093  55.172436   6.979 0.000000527 ***
## TEMP        -8.292527   1.299045  -6.384 0.000002010 ***
## I(TEMP^2)     0.059823   0.007549   7.925 0.000000069 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.376 on 22 degrees of freedom
## Multiple R-squared:  0.9594, Adjusted R-squared:  0.9557
## F-statistic: 259.7 on 2 and 22 DF,  p-value: 4.991e-16
```



Example 5.2 (ctn)

Solution:

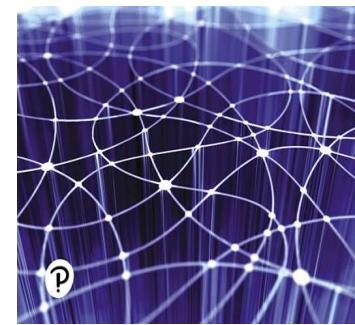
(b) We perform the t-test for the cubic term.

$$H_0: \beta_3 = 0$$

$$H_a: \beta_3 \neq 0$$

From the MINITAB printout on next slide, the p-value is 0.911. Since p-value > $\alpha = .05$, we fail to reject null hypothesis. There is insufficient evidence of a third-order relationship between peak load and high temperature.

Consequently, we will drop the cubic term, $\beta_3 x^3$, from the model.



Example 5.2 (ctn)

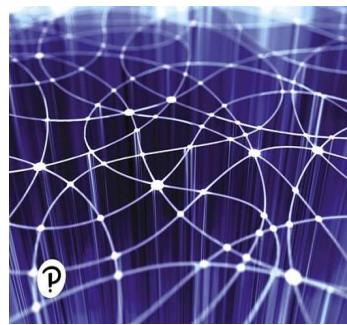
Solution:

(c) We fit the **quadratic model**. If β_2 is positive, then the peak power load y increases at an increasing rate with temperature x . Consequently, we test

$$H_0: \beta_2 = 0$$

$$H_a: \beta_2 > 0$$

From the MINITAB printout on next slide, the two-tailed p-value is 0.000. Since the one-tailed p-value, $p = \frac{0}{2} = 0 < \alpha = .05$, we reject null hypothesis. We conclude the peak power load increases at an increasing rate with temperature.



Example 5.2 (ctn)

Solution:

(d) The prediction equation for the quadratic model is

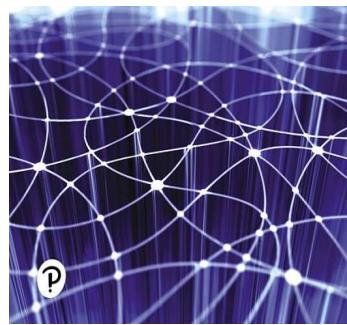
$$\hat{y} = 385 - 8.29x + .0598x^2$$

Then we perform the overall significance F-test

$$H_0: \beta_1 = \beta_2 = 0$$

H_a : At least one of the two
coefficients is nonzero

From the printout in the previous slide, p-value =0 indicates that the model is statistically useful.



Example 5.2 (ctn)

Solution:

(d)

The adjusted- $R^2 = .956$, which imply that more than 95% of the sample variation in peak power loads can be explained by the second-order model.

The standard deviation for the model is $s = 5.376$, which implies the model can predict peak load to within about $2s = 10.75$ megawatts of its true value.

In summary, based on overall F-test, high value of R_a^2 and reasonably small value of $2s$, we recommend using this equation to predict power loads for the company.