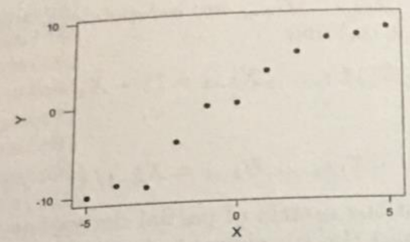
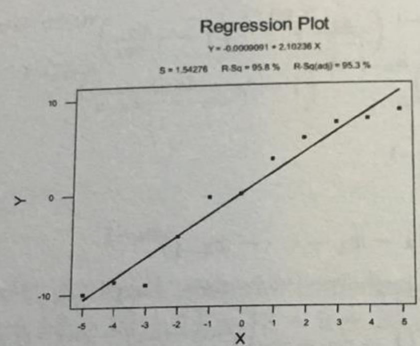
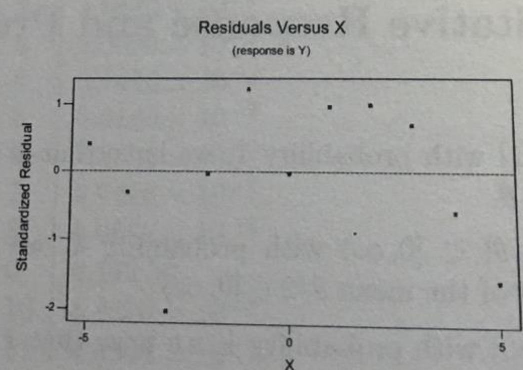
10.3.4(a)



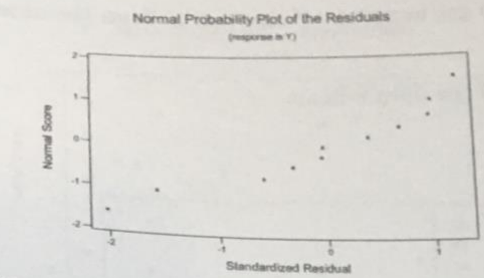
(b) The least-squares estimates of β_1 and β_2 are given by $b_2 = 2.1024$ and $b_1 = \bar{y} = -0.00091$, so the least-squares line is given by $y = -0.00091 + 2.1024x$. A scatter plot of the data together with a plot of the least-squares line follows.



(c) The plot of the standardized residuals against X follows.



(d) A normal probability plot of the standardized residuals is given below.



(e) Both graphs indicate that the normal simple linear regression model is reasonable.
(f) A .95-confidence interval for the intercept is given by

$$-0.00091 \pm 0.4652\,(2.2622) = (-1.0533, 1.0515)$$

and a .95-confidence interval for the slope is given by $2.1024 \pm 0.1471 \cdot 2.2622 = (1.7696, 2.4352)$.
(g) The ANOVA table is follows.

Source	Df	SS	MS
X	1	486.19	486.19
Error	9	21.42	2.38
Total	10	507.61	

The F statistic for testing $H_0 : \beta_2 = 0$ is given by $F = 486.19/2.38 = 204.28$ and, since $F \sim F(1, 9)$ under H_0, the P-value is given by $P(F > 204.28) = .000$, so we reject the null hypothesis of no effect between X and Y.
(h) The proportion of the observed variation in the response that is being explained by changes in the predictor is given by the coefficient of determination $R^2 = 486.19/507.61 = .9578$.
(i) The prediction is given by $y = -0.00091 + 2.1024\,(0) = -0.00091$. This is an interpolation because 0.0 is in the range of observed X values. The standard error of this prediction is, since $\bar{x} = 0$ (using Corollary 10.3.1), $(2.38/11)^{1/2} = 0.46515$.
(j) The prediction is given by $y = -0.00091 + 2.1024\,(6) = 12.613$. This is an extrapolation because 6 is not in the range of observed X values. The standard error of this prediction is, since $\bar{x} = 0$ (using Corollary 10.3.1),

$$(2.38)^{1/2}\left(\frac{1}{11} + \frac{(6-0)^2}{110}\right)^{1/2} = 0.99763.$$

(k) The prediction is given by $y = -0.00091 + 2.1024\,(20) = 42.047$. This is an extrapolation because 12 is not in the range of observed X values. The standard error of this prediction is, since $\bar{x} = 0$ (using Corollary 10.3.1),

$$(2.38)^{1/2}\left(\frac{1}{11} + \frac{(20-0)^2}{110}\right)^{1/2} = 2.9784.$$

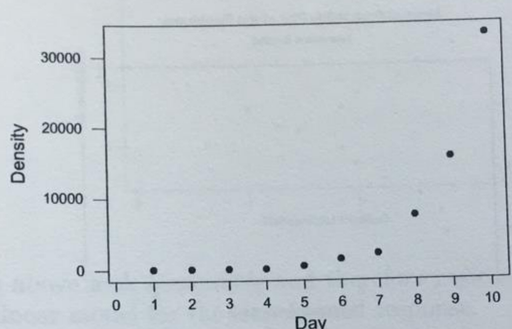
The standard errors get larger as we move away from the observed X values.

10.3.5

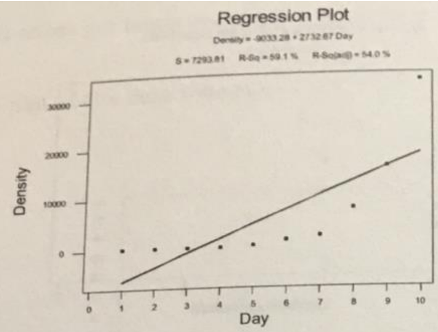
(a) A scatter plot of the data follows.

10.3.6

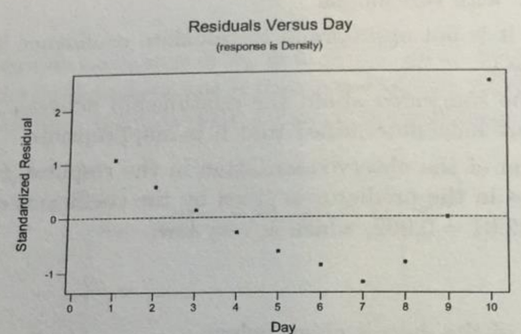
(a) A scatter plot of the data is given below.



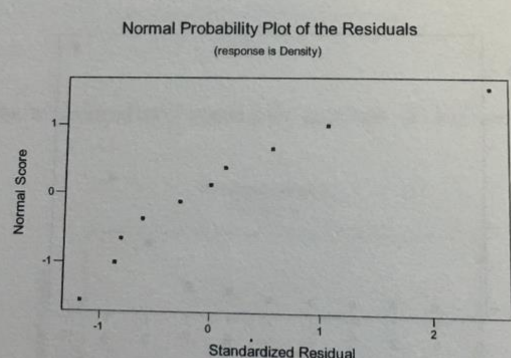
(b) The least-squares estimates of β_1 and β_2 are given by $b_2 = 2732.67$ and $b_1 = -9033.28$, respectively. The least-squares line is then given by $y = -9033.28 + 2732.67x$. A scatter plot of the data together with a plot of the least-squares line follows.



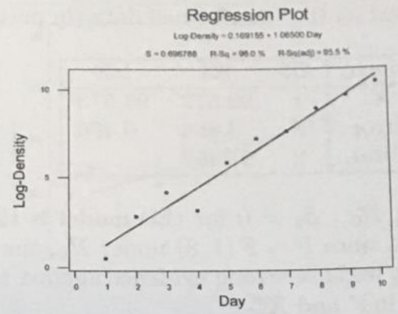
(c) A plot of the standardized residuals against X follows.



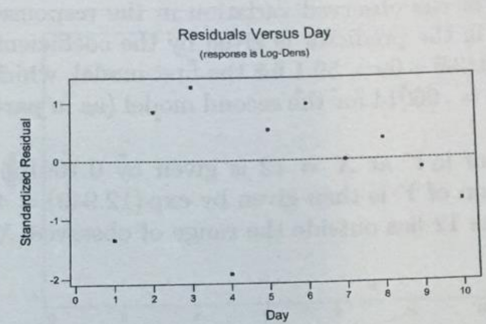
(d) A normal probability plot of the standardized residuals follows.



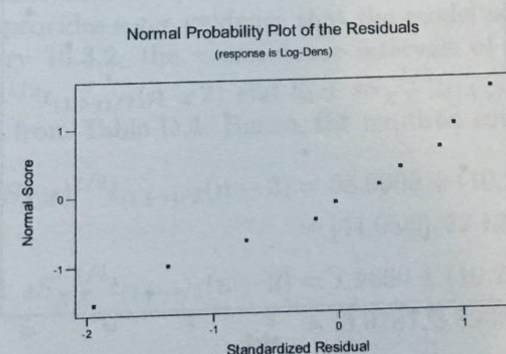
(e) The plot of the standardized residuals against X indicates very clearly that there is a problem with this model.
(f) Taking the logarithm of the response, we obtain the least-squares line given by $\ln(y) = 0.169155 + 1.06500x$. A scatter plot of the data together with the least-squares line follows



A plot of the standardized residuals against X follows.



A normal probability plot of the standardized residuals follows.



Both graphs above look reasonable and therefore indicate no evidence against the normal linear model for the transformed response.

(g) As we can see from the scatter plot in part (a), the relationship between X and Y is definitely non-linear, and therefore it is not appropriate to calculate confidence intervals for the intercept and slope. However, after transforming the response, the relationship looks quite linear, so for this model 0.95-confidence intervals for the intercept and the slope are given by $0.169155 \pm 0.4760\,(2.306) = (-.9285, 1.2668)$ and $1.065 \pm 0.07671\,(2.306) = (.88811, 1.2419)$, respectively.

(h) The ANOVA table based on the transformed data (in part f) is given below.

Source	Df	SS	MS
X	1	93.573	93.573
Error	8	3.884	0.486
Total	9	97.458	

The F statistic for testing $H_0 : \beta_2 = 0$ for this model is then given by $F = 93.573/3.884 = 24.092$ and, since $F \sim F(1, 8)$ under H_0, the P-value is $P(F > 24.092) = 0.000$. Therefore, we have strong evidence against the null hypothesis of no relationship between $\ln Y$ and X.

(i) Yes, we can conclude that there is a relationship. We can then express the relationship between X and Y as $E(\ln Y \mid X = x) = 0.169155 + 1.06500x$.

(j) The proportion of the observed variation in the response that is being explained by changes in the predictor is given by the coefficient of determination $R^2 = 616068769/1.042E+09 = 59.1$ for the first model, which is quite low, and $R^2 = 93.573/97.458 = .96014$ for the second model (as in part f), which is quite high.

(k) The prediction of $\ln Y$ at $X = 12$ is given by $0.169155 + 1.06500(12) = 12.949$. The prediction of Y is then given by $\exp(12.949) = 4.2042 \times 10^5$. This is an extrapolation as 12 lies outside the range of observed X values.

10.3.8
(a) From the relationship, $Z = Y - E(Y|X)$ and

$$E(Z|X) = E(Y - E(Y|X)|X) = E(Y|X) - E(Y|X) = 0.$$

(b) The covariance can be written as

$$\text{Cov}(E(Y \mid X), Z) = E(E(Y \mid X)Z) - E(E(Y \mid X))E(Z).$$

Theorem 3.5.2 implies $E(Z) = E(E(Y \mid X))$ and $E(Z \mid X) = 0$ from part (a). So, $E(Z) = E(E(Z \mid X)) = E(0) = 0$. In a similar vein, $E(E(Y \mid X)Z) = E(E(E(Y \mid X)Z|X))$ and $E(E(Y \mid X)Z|X) = E(Y \mid X)E(Z|X) = 0$. Therefore, $\text{Cov}(E(Y \mid X), Z) = 0 - 0 = 0$.
(c) Given $X = x$, $E(Y|X = x)$ is constant. So, the conditional cdf of Y given $X = x$ is

$$\begin{aligned} F_{Y|X}(y \mid x) &= P(Y \leq y \mid x) = P(Y - E(Y \mid X = x) \leq y - E(Y \mid X = x) \mid x) \\ &= P(Z \leq y - E(Y \mid X = x) \mid x) = F_Z(y - E(Y \mid X = x)). \end{aligned}$$

We see from this that the conditional distribution Y given X depends on X only through its conditional mean $E(Y \mid X)$.

10.3.9 In general, $E(Y \mid X) = \exp(\beta_1 + \beta_2 X)$ is not a simple linear regression model since it cannot be written in the form $E(Y \mid X) = \beta_1^* + \beta_2^* V$ where V is an observed variable and the β_i^* are unobserved parameter values.

10.3.10 Corollary 3.6.1 implies that

$$Y = E(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - E(X)).$$

By letting $\beta_2 = \text{Cov}(X,Y)/\text{Var}(X)$ and $\beta_1 = E(Y) - \beta_2 E(X)$, the model becomes $Y = \beta_1 + \beta_2 X$. Hence, it is a simple linear regression model where $Z \equiv 0$.