**3.3** Salary vs. GPA

a. We consider Figure 3.16. From this figure, we find the regression equation to be,

$$Y = 14.8156 + 5.70657X$$

Thus, we see that the relevant coefficients are,

$$b_0 = 14.8156$$
$$b_1 = 5.70657$$

This value of $b_0$ is equal to the initial salary value for a student with a GPA of 0. In this sense, it does not make sense. But, if we assume that this model is only used for students that have passed, i.e. $GPA \neq 0$, this begins to make more sense. In this way, we can call $b_0$ a base salary that is then modified by higher GPA's

b. Using the equation from Figure 3.16, we compute,

$$Y = 14.8156 + 5.70657 * 3.25 = 33.362$$

Because a GPA of 3.25 is within our experimental region, we note that this prediction of salary is also a point estimate for the value as well.

c. We finally compute the Coefficients using the following equations,

$$SS_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$
$$SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Using R, we compute the values to be exactly as proposed in the figure.

**3.7** Copier Data

a. Using the figure, we identify the model to be,

$$Y = 11.4641 + 24.6022X$$

Thus, the relevant coefficients are,

$$b_0 = 11.4641$$
$$b_1 = 24.6022$$

This value of $b_0$ corresponds to the initial amount of time taken to service 0 copiers. In this sense, this value doesn't make sense. If we make a few assumptions regarding how this job is performed, we can begin to extrapolate more value. For instance if we assume that the fixing of the copiers takes place in one location, we can call this value an amount of time required to set up for working, or aquiring more parts.

b. Using the equation from Figure 3.18, we compute,

$$Y = 11.4641 + 24.6022 * 4 = 109.8729$$

Because 4 copiers is within our experimental regfion, we note that this prediction of time is also a point estimate as well.

c. Finally, we compute the coefficients using the following equations,

$$SS_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$
$$SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Using R, we compute the values to be exactly as proposed in the figure.

**3.9** Enterprise Industries

Using the scatter plot in Figure 3.19, we begin making some superficial remarks. First, we note that the overall shape of the data is roughly linear. This is a good indicator that a linear model will fit the data well. Next, we consider the nature of the relationship between $y$ and $x_4$. $x_4$ is the difference between the average price of detergent in the market and the cost of Fresh brand detergent. Hence, $x_4 > 0 \Rightarrow$ The average cost is greater than the cost of Fresh, and $x_4 < 0 \Rightarrow$ the averate cost is less than the cost of fresh. So, we may begin to establish a linear relationship between the data. Finally, we note that the data is scattered about the linear trend, which also implies that the data may be described by a least squares regression.

**3.20** MINITAB

a. We shall report a point estimate and a 95% CI for the mean starting salary for a student with a GPA of 3.25. To begin, we consider the interval equation,

$$I = \hat{y} \pm t_{\alpha/2}^{n-2} s\sqrt{Distance}$$

First, we compte the distance value as,

$$Distance = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}} = \frac{1}{7} + \frac{(3.25 - 3.0814)^2}{1.840686} = .158295$$

From the Minitab output, we know $s = .536321$ and we know that $t_{.025}^{5} = 2.571$ Hence, we arrive at the interval,

$$I = (32.81339, 33.91061)$$

As reported by Minitab.

b. We shall now report a point prediction and interval for an individual with a GPA of 3.25. We consider the interval equation for this situation,

$$I = \hat{y} \pm t_{\alpha/2}^{n-2} s\sqrt{1 + Distance}$$

We note that this equation is nearly identical to the one above, with only the addition of the 1 under the root. Thus, we may compute this interval as above to find,

$$I = (31.87799, 34.84601)$$

This too agrees with the interval generated in the minitab output.

**3.21** SAS

a. From the SAS output, we see that the CI generated from the data for 4 copiers being serviced is,

$$I = (106.7207, 113.0252)$$

Hence, the company could say that they are 95% confident that the mean of all times for their employees to service 4 copiers falls within this interval.

b. Similarly, from the SAS output, we see that the CI generated from the data for fixing four copiers on a single call is,

$$I = (98.9671, 120.7788)$$

Hence the company could say that they are 95% confident that for a future call, the amount of time for an employee to fix 4 copiers will fall within this interval.

c. Judging from the interval computed for the mean time for fixing four copiers, we know that we have predicted the mean to be in the range $(106.7207, 113.0252)$. Allowing for the largest possible amount of time for the service call, the company could reasonably allocate 113 minutes to make the call for fixing 4 copiers. It is also intersting to note that this value is approximately the time allocated for fixing five copiers less $b_1$, the predicted slope of the linear model.

**3.35** $\frac{1}{x}$ model

Consider the data from the problem,

| Time | 8.0 | 4.7 | 3.7 | 2.8 | 8.9 | 5.8 | 2.0 | 1.9 | 3.3 |
|---|---|---|---|---|---|---|---|---|---|
| Experience | 1 | 8 | 4 | 16 | 1 | 2 | 12 | 5 | 3 |

Because our model is related to the values of $\frac{1}{x}$, we will also compute,

| Time | 8.0 | 4.7 | 3.7 | 2.8 | 8.9 | 5.8 | 2.0 | 1.9 | 3.3 |
|---|---|---|---|---|---|---|---|---|---|
| Experience | 1 | 8 | 4 | 16 | 1 | 2 | 12 | 5 | 3 |
| $\frac{1}{experience}$ | 1 | 0.125 | 0.25 | 0.0625 | 1 | 0.5 | 0.0833 | 0.2 | 0.333 |

Next, we shall compute the linear model using $R$,

$$\hat{y} = 6.3537x + 2.0575$$

Thus, we may compute the point prediction for $x = 5 \Rightarrow \frac{1}{x} = \frac{1}{5}$,

$$\hat{y} = \frac{6.3537}{5} + 2.0757 = 3.32824$$

Next, we begin the computations of our prediction interval using,

$$I = \hat{y} \pm t_{\alpha/2}^{n-2} s\sqrt{1 + Distance} = \hat{y} \pm \Delta$$

First, we compute $t_{\alpha/2}^{n-2} = t_{.025}^{7} = 2.365$ from the table.
Next, we compute $s$ as,

$$\begin{aligned} s &= \sqrt{\frac{SSE}{n-2}} \\ &= \sqrt{\frac{\sum y_i^2 - \left(b_0 \sum y_i + b_1 \sum x_i y_i\right)}{7}} \\ &= \sqrt{\frac{7.41919525}{7}} \\ &= 1.029507 \end{aligned}$$

Finally, we compute $Distance$ as,

$$\begin{aligned} Distance &= \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}} \\ &= \frac{1}{9} + \frac{(.2 - .39491)^2}{1.08652} && = .14607493 \end{aligned}$$

Then

$$\Delta = 2.365 * 1.029507 * \sqrt{1 + 0.146074} = 2.6065558$$

Thus,

$$I = (0.7216842, 5.934796)$$

As desired.