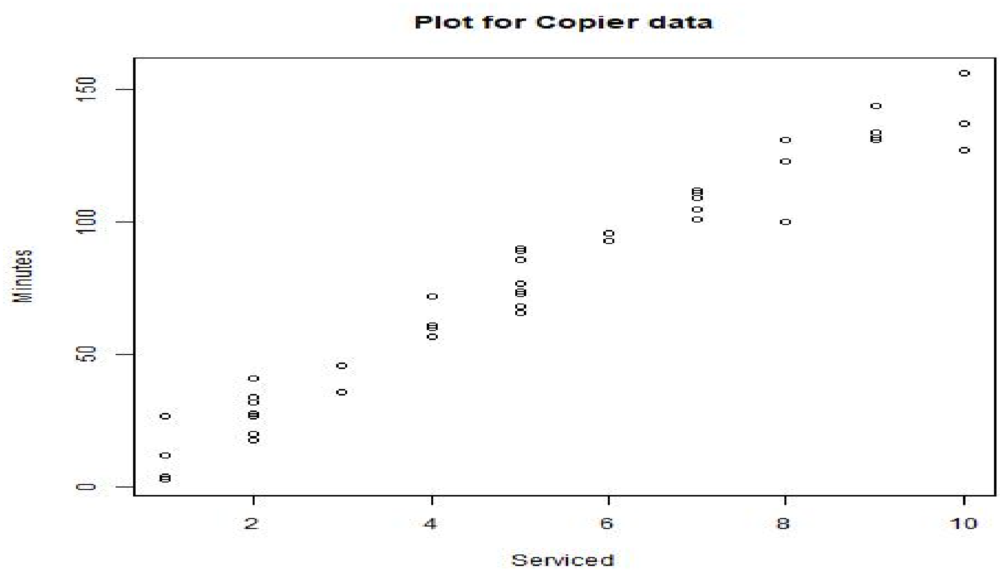
Stat 6021: Homework Set 2 Solutions

1. (a) The scatterplot is shown below. We can see there is a strong positive linear association between the total time spent by the service person and the number of copiers serviced.



   (b) The correlation is 0.978517. This indicates a strong positive linear relationship.

   (c) Since the scatterplot shows a linear relationship, we can reliably interpret the correlation as a measure of linear relation.

   (d) The 95% CI for the slope is (14.061010, 16.009486).

```
> confint(result, level=0.95)
                 2.5 %    97.5 %
(Intercept) -6.234843  5.074529
Serviced    14.061010 16.009486
```

   (e) The 95% PI for the total service time for a service person who services 5 copiers is (56.42133, 92.77084) minutes.

```
newdata <- data.frame(Serviced=5)
predict(result, newdata, interval="prediction")
```

(f) The residual for the first observation is -9.490339 minutes. This means the total service time for the first service person is 9.490339 minutes shorter than his/her predicted total service time based on our regression equation.

(g) The mean of the residuals is $-2.612204 \times 10^{-16}$. We know that the mean residual is 0, theoretically, for OLS. This value very closely matches with the theory.

2. (a) $H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$.

   The t statistic is $\frac{\hat{\beta}_1 - 0}{s\{\hat{\beta}_1\}} = \frac{4}{0.4690} = 8.528$.

   The p-value for testing $\beta_1 = 0$ is $2 \times (1 - pt(8.528, 8))$ which is about 0. Thus, we can reject the null hypothesis. Our data suggests there is a linear relationship between number of broken ampules and number of transfers.

   Alternatively, the critical value is $qt(0.975, 8)$ which is 2.3060. Since the t statistic is greater than the critical value, we can reject the null hypothesis. Our data suggests there is a linear relationship between number of broken ampules and number of transfers.

   (b) $\hat{\beta}_1 \pm t_{0.975,8} s\{\hat{\beta}_1\} = 4 \pm 2.306004 \times 0.4690 = (2.918484, 5.081516)$.

   (c) $H_0 : \beta_0 = 9$.

   $H_0 : \beta_0 \neq 9$.

   Test statistic $= \frac{\hat{\beta}_0 - 9}{s\{\hat{\beta}_0\}} = \frac{10.2-9}{0.6633} = 1.809$. The p-value is $2 \times (1 - pt(1.809, 8)) = 0.108$. We would fail to reject the null. The data does not support his belief.

   Alternatively, the critical value is $qt(0.975, 8)$ which is 2.3060. Since the t statistic is less than the critical value, we fail to reject the null hypothesis. The data does not support his belief.

   (d) When $X = 2$, $\hat{y} = 10.2 + 4(2) = 18.2$.

   95% CI for μ_0:

$$
\begin{aligned}
\hat{\mu}_0 \pm t_{0.975,8} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} &= 18.2 \pm 2.306004 \times 1.483 \sqrt{\frac{1}{10} + \frac{(2-1)^2}{10}} \\
&= (16.67062, 19.72938)
\end{aligned}
$$

   95% PI for new Y:

$$
\begin{aligned}
\hat{y}_0 \pm t_{0.975,8} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} &= 18.2 \pm 2.306004 \times 1.483 \sqrt{1 + \frac{1}{10} + \frac{(2-1)^2}{10}} \\
&= (14.45379, 21.94621)
\end{aligned}
$$

   (e) When number of transfers is 1, the intervals from the previous part get narrower, since we are computing intervals when x_0 is closer to the mean.

   (f) $F = \frac{\text{MSR}}{\text{MSres}} = \frac{160}{2.2} = 72.72727$.

(g) $R^2 = \dfrac{\text{SSR}}{\text{SST}} = \frac{160}{160+17.6} = 0.9009009$. This means that about 90% of the variation in number of broken ampules is explained by the number of transfers.

3. (a) The plot will be a straight horizontal line.

   (b) When the slope is 0, this means that the value of Y stays the same regardless of the value of X. In other words, when the value of X changes, the value of Y does not change. This means the variables are not linearly related.