University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Omitting a point in simple regression - effect on $\hat{\beta}_1$ and $\hat{\beta}_0$

In this document we will explore the effect of deleting a single point in simple regression.

Notation:

1. $S_{XY} = \sum_{i=1}^n (x_i - \bar{x})(y_1 - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n}(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)$.
2. $S_{XX} = \sum_{i=1}^n (x_i - \bar{x})^2$.
3. $S_{YY} = \sum_{i=1}^n (y_1 - \bar{y})^2$.
4. In the discussion below, the symbol $|i$ in a subscript means that the quantity is calculated without point i. For example, $\bar{y}_{|i}$ is the sample mean of $n-1$ values of y (all but y_i). Another example, $S_{XY|i}$ is the quantity S_{XY} after the point (x_i, y_i) is omitted.

Questions:

a. Give an expression of $\bar{x}_{|i}$ in terms of $\bar{x}$ and x_i.
SOLUTION:

$$\begin{aligned}(n-1)\bar{x}_{|i} + x_i &= n\bar{x} \\ \bar{x}_{|i} &= \frac{n}{n-1}\bar{x} - \frac{1}{n-1}x_i\end{aligned}$$

b. Give an expression of $\bar{y}_{|i}$ in terms of $\bar{y}$ and y_i.
SOLUTION:

$$\begin{aligned}(n-1)\bar{y}_{|i} + y_i &= n\bar{y} \\ \bar{y}_{|i} &= \frac{n}{n-1}\bar{y} - \frac{1}{n-1}y_i\end{aligned}$$

c. Show that $S_{XY|i} = S_{XY} - \frac{n}{n-1}(x_i - \bar{x})(y_1 - \bar{y})$.
SOLUTION:

$$\begin{aligned}S_{XY} &= \sum_{j=1}^n x_j y_j - \frac{1}{n}\left(\sum_{j=1}^n x_j\right)\left(\sum_{j=1}^n y_j\right) \\ &= \sum_{j\neq i}^n x_j y_j + x_i y_i - \frac{1}{n}\left(\sum_{j\neq i}^n x_j + x_i\right)\left(\sum_{j\neq i}^n y_j + y_i\right) \\ &= \sum_{j\neq i}^n x_j y_j - \frac{1}{n-1}\left(\sum_{j\neq i}^n x_j + x_i\right)\left(\sum_{j\neq i}^n y_j + y_i\right) \\ &+ \quad x_i y_i + \frac{1}{n(n-1)}\left(\sum_{j\neq i}^n x_j + x_i\right)\left(\sum_{j\neq i}^n y_j + y_i\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\neq i}^{n} x_j y_j - \frac{1}{n-1}\left(\sum_{j\neq i}^{n} x_j\right)\left(\sum_{j\neq i}^{n} y_j\right) \\
&- \frac{1}{n-1} y_i \left(\sum_{j\neq i}^{n} x_j\right) - \frac{1}{n-1} x_i \left(\sum_{j\neq i}^{n} y_j\right) - \frac{1}{n-1} x_i y_i \\
&+ x_i y_i + \frac{1}{n(n-1)}\left(\sum_{j\neq i}^{n} x_j + x_i\right)\left(\sum_{j\neq i}^{n} y_j + y_i\right) \\
&= S_{XY|i} \\
&- \frac{1}{n-1} y_i \left(n\bar{x} - x_i\right) - \frac{1}{n-1} x_i \left(n\bar{y} - y_i\right) - \frac{1}{n-1} x_i y_i + x_i y_i + \frac{1}{n(n-1)} \left(n\bar{x}\right)\left(n\bar{y}\right) \\
&= S_{XY|i} + \frac{1}{n-1}\left(-n\bar{x} y_i + x_i y_i - n\bar{y} x_i + x_i y_i - x_i y_i + (n-1) x_i y_i + n\bar{x}\bar{y}\right) \\
&= S_{XY|i} + \frac{1}{n-1}\left(-n\bar{x} y_i + n x_i y_i - n\bar{y} x_i + n\bar{x}\bar{y}\right) \\
&= S_{XY|i} + \frac{n}{n-1}\left[(x_i - \bar{x})(y_i - \bar{y})\right].
\end{aligned}
$$

Therefore,

$$S_{XY|i} = S_{XY} - \frac{n}{n-1}(x_i - \bar{x})(y_1 - \bar{y}).$$

d. Similarly to (c) we can show that

$$S_{XX|i} = S_{XX} - \frac{n}{n-1}(x_i - \bar{x})^2$$

and

$$S_{YY|i} = S_{yy} - \frac{n}{n-1}(y_i - \bar{y})^2$$

Use the results above to find expressions for $\hat{\beta}_{1|i}$ and $\beta_{0|i}$. In other words, find expressions for $\hat{\beta}_1$ and $\hat{\beta}_0$ after point i was deleted.