Appendix A

Finding least squares estimators for simple linear regression.

The regression line is an equation of the form

$$y' = a + bx \ .$$

Here, x is an observed value of the independent variable, y is an observed value of the dependent variable, y' is the estimated value of y, a is the intercept of the regression line and b is the slope of the regression line. What we would like to do is find estimators of a and b which will give us a best line. Best is usually defined as the least squares line. The least squares line is the line which will minimize the sum of the squared differences between y and y'. The sum of the squared differences is

$$\Sigma(y - y')^2,$$

which can also be written as

$$\Sigma(y - a - bx)^2.$$

We want to minimize this sum of squares. This is where the name for the "least squares approach" comes from.

We are used to thinking of this quantity to be minimized,

$$\Sigma(y - a - bx)^2,$$

as a function of x and y, but we can see this last version can also be viewed as a function of a and b. We have then

$$f(a,b) = \Sigma(y - a - bx)^2.$$

Remember that the sigma notation simply indicates summation and that the derivative of a sum is the sum of the derivatives. We can thus take derivatives inside the summation and they will add appropriately:

$$\frac{\partial f}{\partial a} = \Sigma\ 2(y - a - bx)^1\ (-1)$$

$$\frac{\partial f}{\partial b} = \Sigma\ 2(y - a - bx)^1\ (-x).$$

If we set these two derivatives equal to zero and solve the resulting two equations simultaneously, we find:

$\Sigma\ 2(y - a - bx)^1\ (-1) = 0$	Partial with respect to a.
$\Sigma\ 2(y - a - bx)^1\ (-x) = 0$	Partial with respect to b.
$-2\Sigma\ (y - a - bx) = 0$	Move constants outside Σ.
$-2\Sigma\ (y - a - bx)\ (x) = 0$	Move constants outside Σ.
$\Sigma\ (y - a - bx) = 0$	Divide by -2.
$\Sigma\ (y - a - bx)\ (x) = 0$	Divide by -2.
$\Sigma y - \Sigma a - \Sigma bx = 0$	Distribute Σ.
$\Sigma\ (xy - ax - bx^2) = 0$	Multiply through by x.
$\Sigma y - na - \Sigma bx = 0$	Summation rule.
$\Sigma xy - \Sigma ax - \Sigma bx^2 = 0$	Distribute Σ.

Equation 1:	$\Sigma y - na - b\Sigma x = 0$	Move constant outside Σ.
Equation 2:	$\Sigma xy - a\Sigma x - b\Sigma x^2 = 0$	Move constant outside Σ.

We now solve the Equation 1 for a and substitute that result for a in Equation 2 and solve the resulting equation.

$\Sigma y - na - b\Sigma x = 0$	Equation 1.
$na = \Sigma y - b\Sigma x$	Move terms across equal sign.
$a = \dfrac{(\Sigma y - b\Sigma x)}{n}$	Divide by n.

Substituting this result for a into Equation 2 we find:

$\Sigma xy - a\Sigma x - b\Sigma x^2 = 0$	Equation 2.
$\Sigma xy - \dfrac{(\Sigma y - b\Sigma x)\Sigma x}{n} - b\Sigma x^2 = 0$	Substitution.
$\Sigma xy - \underline{\Sigma x\Sigma y - b\Sigma x\Sigma x} - b\Sigma x^2 = 0$	Remove parentheses.

n

$$\Sigma xy - \frac{\Sigma x \Sigma y}{n} - \frac{b(\Sigma x)^2}{n} - b\Sigma x^2 = 0$$ Separate fraction into two fractions

$$-\frac{b(\Sigma x)^2}{n} + b\Sigma x^2 = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$ Move terms across equal sign.

$$b\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n} \right) = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$ Factor b out of two terms.

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$ Divide both sides by $\Sigma x^2 - \frac{(\Sigma x)^2}{n}$

This is the solution for the slope, b. To find the solution for a w can take equation 1 and solve for a.

$$\Sigma y - na - b\Sigma x = 0$$ Equation 1

$$\frac{\Sigma y}{n} - a - \frac{b\Sigma x}{n} = 0$$ Divide by n.

$$a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$$ Move terms across equal sign.

Recognizing the formulae for $\bar{y}$ and $\bar{x}$, we can rewrite the last equatio as:

$$a = \bar{y} - b\bar{x} \; .$$