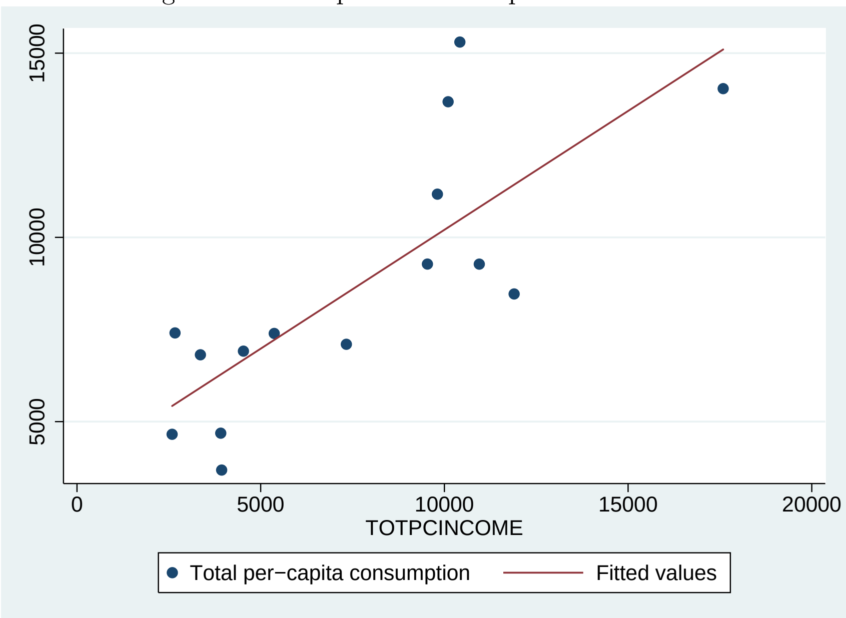
Calculation of β's parameters

Consider individual data

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad \text{where} \quad i = 1, 2, ...n \tag{3}$$

Figure 1: Scatterplot of consumption and income



Now, recalling that

$$E(u) = 0 \tag{4}$$

then

$$Cov(x, u) = E(xu) = 0 \tag{5}$$

and by substitution

$$E(y - \beta_0 - \beta_1 x) = 0 \tag{6}$$

$$E[x(y - \beta_0 - \beta_1 x)] = 0 \tag{7}$$

Now taking the sample values of 6 and 7

$$n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{8}$$

$$n^{-1} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{9}$$

Here we applied the *method of moments*, but other approaches could also be used. Now, let's solve 8 and 9 by $\hat{\beta}_0$ and $\hat{\beta}_1$.

Equation 8 can be expressed as:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \tag{10}$$

and then

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{11}$$

This is the **sample intercept**.

To get $\hat{\beta}_1$, let's substitute the expression of $\hat{\beta}_0$ from 11 into 9 (we can drop n^{-1} as it's only a rescaling parameter which doesn't affect the slope). After some algebraic transformations, using summation properties, and given that:

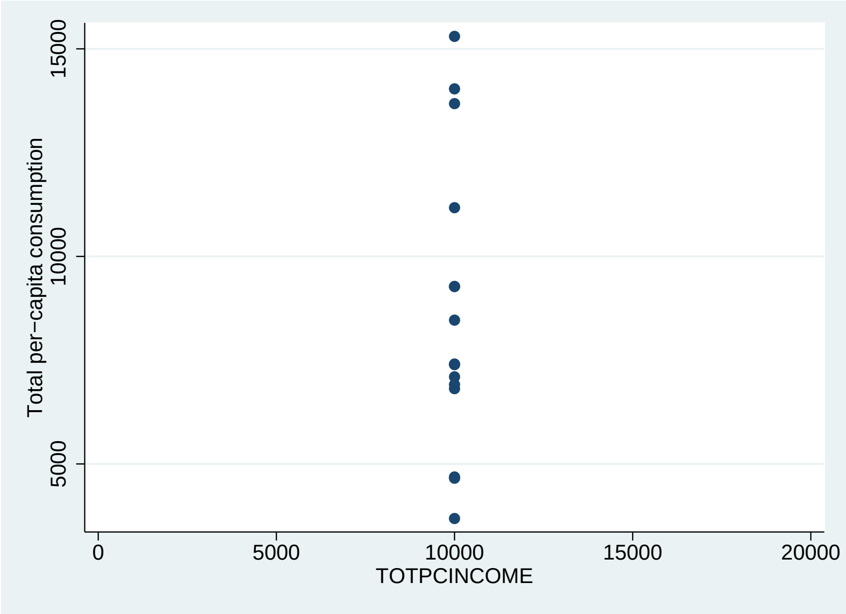
$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0 \tag{12}$$

we get

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{13}$$

the **estimated slope**, that is the sample counterpart of the covariance between x and y divided by the variance of x. If x and y are positively correlated, then $\hat{\beta}_1$ is positive, if they are negatively correlated $\hat{\beta}_1$ will be negative.

Figure 2: Scatterplot of consumption and income when income is fixed



The expressions 11 and 13 are the **Ordinary Least Squares (OLS)** estimates of β_0 and β_1. To get a particular **fitted** value of y when $x = x_i$:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{14}$$

For each value $y = y_i$ there is a fitted value in 14. The distance between the fitted $\hat{y}$ and the actual value y_i is the **residual**, expressed by:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \tag{15}$$

Are $\hat{u}_i$ the same as u_i in equation 3?

Once we get $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\sum_{i=1}^{n} \hat{u_i^2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{16}$$

There are particular values of $\hat{\beta}_0$ and $\hat{\beta}_1$ which minimize 16, the **sum of squared residuals**.

`Quite coincidentally`, these values are given by 8 and 9 (apart from n^{-1}). Instead of the methods of moments, we could have applied a minimization of 16, deriving the **First Order Conditions (FOC)** for the OLS, whose solutions are 11 and 13. This is why they are called "ordinary least squares".

Now we get everything for the regression line of graph 1:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x. \tag{17}$$

This is also called the **sample regression function (SRF)**, why?

Because it is the sample counterpart of equation 2. With a different sample, we would get different values of $\hat{\beta}_0$ and $\hat{\beta}_1$ in equation 17.

Economists always focus on $\hat{\beta}_1$, as they are interested in the change in $\hat{y}$ for a unit change in x. In fact, being

$$\hat{\beta}_1 = \Delta\hat{y}/\Delta x \tag{18}$$

then

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x \tag{19}$$

Let's look at an empirical example.

Figure 3: Scatterplot of consumption and income, sample and population