HG
04.02.2013

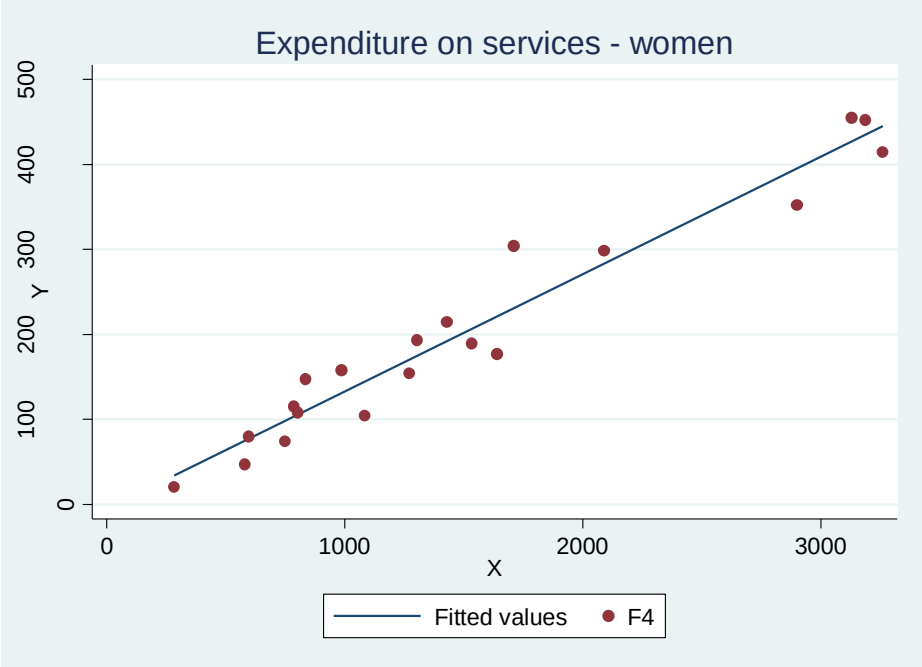
Data: Based on a Hong Kong survey of household expenditure and give **the expenditure of 20 single men (M) and 20 single women (W) on four commodity groups.** In Hong Kong dollars (HKD).

(Data on the net in a Stata file)

		Response variables (Y)	
		Women	**Men**
1.	Housing, including fuel and light	F1	M1
2.	Foodstuffs, including alcohol and tobacco	F2	M2
3.	Other goods, including clothing, footwear and durable goods	F3	M3
4.	Services, including transport and vehicles	F4	M4
		Sum XF	XM

Explanatory variable is income (X) predicted by

XK = total expenditure women, XM = total expenditure men



Stata command: twoway (scatter F4 XF) (lfit F4 XF)

Table 1. Expenditure on services (Y) and total expenditure (X), women

Exp. Y	154	20	455	115	104	193	214	80	352	414
Total exp X	1271	284	3128	786	1084	1303	1428	596	2899	3258
Exp. Y	47	452	108	189	298	158	304	74	147	177
Total exp X	581	3186	804	1533	2088	986	1709	748	836	1639

Model 1 - Fixed-X model:

1) Assume income measured without error by X
2) Assume the $n = 20$ incomes, $x_1, x_2, \ldots, x_n$, are fixed numbers (non random)
3) Assume expenditures $Y_1, Y_2, \ldots, Y_n$ independent and normally distributed

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

with $E(Y_i) = \alpha + \beta x_i \quad (\Leftrightarrow E(\varepsilon_i) = 0)$ and constant variance, $\text{var}(Y_i) = \text{var}(\varepsilon_i) = \sigma^2$ for $i = 1, 2, \ldots, n$.

OLS estimators: $\hat{\beta} = \dfrac{s_{xY}}{s_x^2} = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})\left(Y_i - \bar{Y}\right)}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ unbiased, $\quad \text{var}(\hat{\beta}) = \dfrac{\sigma^2}{(n-1)s_x^2}$

Estimate $\hat{\beta}_{obs} = 0.1383$

$\hat{\alpha} = \bar{Y} - \hat{\beta} \cdot \bar{x}$, unbiased, $\quad \text{var}(\hat{\alpha}) = \sigma^2 \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{(n-1)s_x^2} \right)$

Estimate $\hat{\alpha}_{obs} = -5.7345$

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \frac{n-1}{n-2}\left(S_y^2 - \hat{\beta}^2 s_x^2\right), \text{ unbiased, where}$$

$\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta} x_i$, $i = 1, 2, \text{K}, n$ are the residuals

Estimate $\hat{\sigma}^2_{obs} = 1013.7204$

Theorem: $T = \dfrac{\hat{\beta} - \beta}{SE(\hat{\beta})} \sim t_{n-2}$ – i.e., t-distributed with $n-2$ degrees of freedom for any β and fixed values, $x_1, x_2, \ldots, x_n$, and where $SE(\hat{\beta}) = \sqrt{\hat{\text{var}}(\hat{\beta})} = \dfrac{\hat{\sigma}}{\sqrt{(n-1)s_x^2}}$

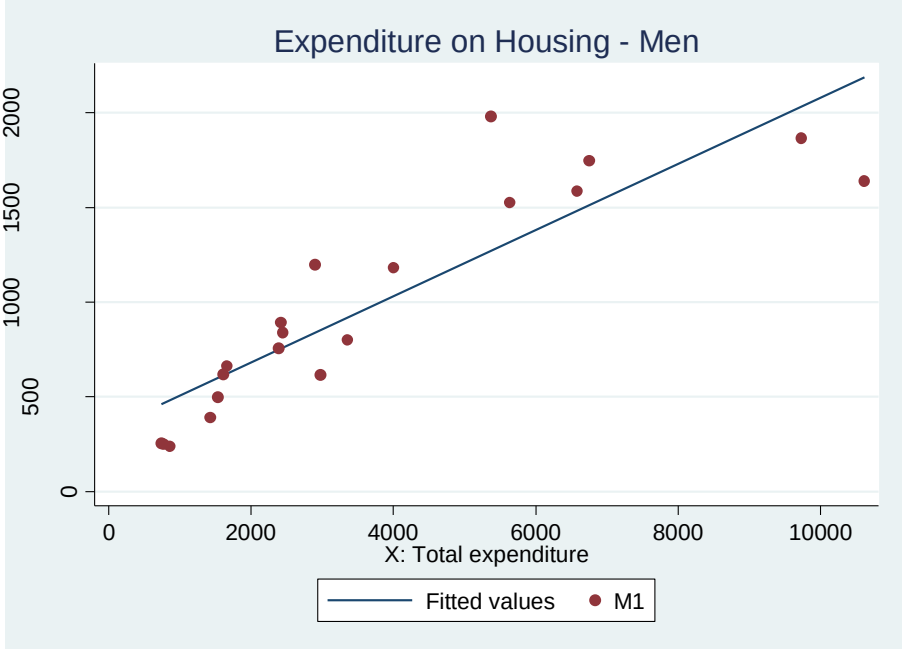
From this we get (e.g.) a **95% confidence interval (CI)** for β, (L, U), satisfying

$$P(L \le \beta \le U) = 0.95$$

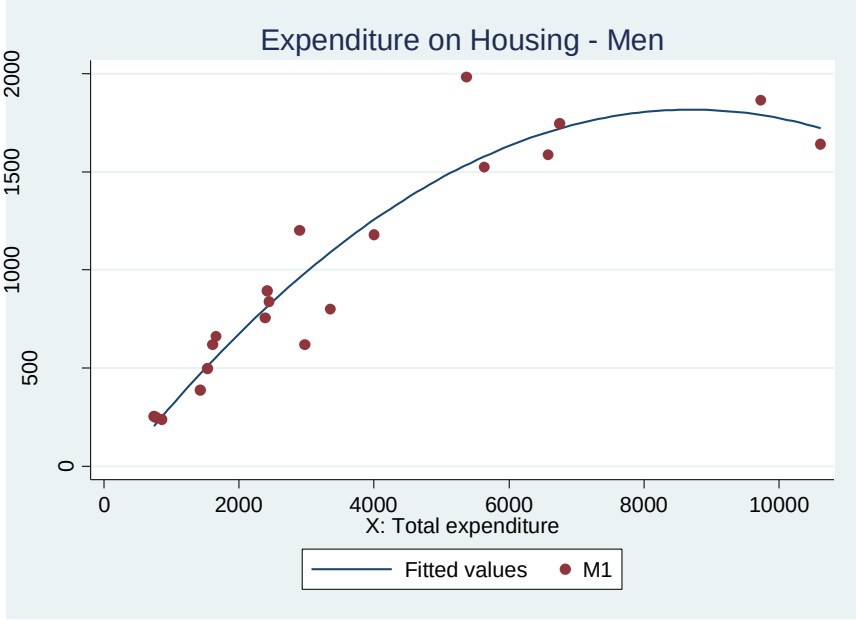
where $L = \hat{\beta} - t_{n-2,\,0.025} \cdot SE(\hat{\beta})$, $\quad U = \hat{\beta} + t_{n-2,\,0.025} \cdot SE(\hat{\beta})$, and $t_{n-2,\,0.025}$ is the upper 2.5% point in the t_{n-2} - distribution (for example Løvås gives; $t_{18,\,0.025} = 1.734$).

Relation of X on expenditure on housing for men.

Linear and homoscedastic regression model, $\mu(x) = E(Y \mid x) = \alpha + \beta x$

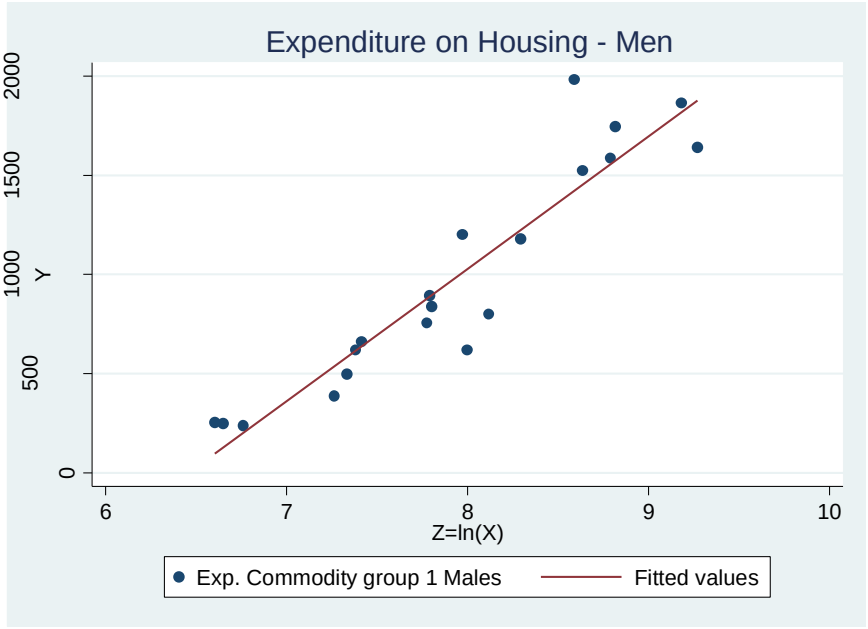


Quadratic (and homoscedastic) regression model, $\mu(x) = E(Y \mid X = x) = \beta_0 + \beta_1 x + \beta_2 x^2$



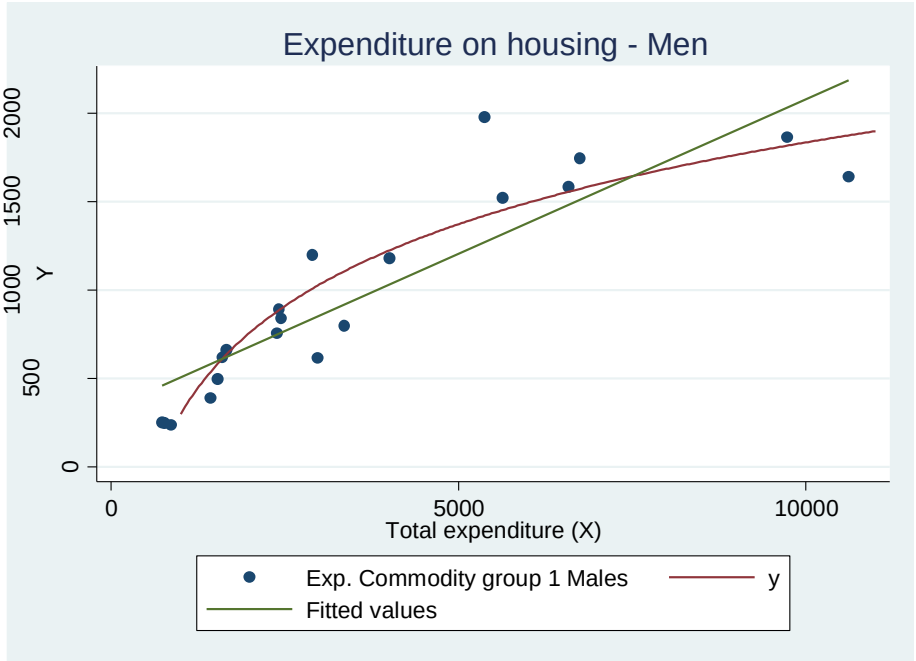
Stata command: twoway (scatter M1 XM) (qfit M1 XM)

Model based on log-transformed X, $E(Y \mid Z = z) = \alpha + \beta z = \mu(z)$, where $Z = \ln(X)$



Stata command: twoway (scatter M1 LXM) (lfit M1 LXM)

The fitted model on original X-scale: $E(Y \mid X = x) = \alpha + \beta \ln(x) = \mu(\ln(x))$



Stata command:
twoway (scatter M1 XM) (function y=-4314.101 +667.7628*ln(x), range(1000 11000)) (lfit M1 XM)